

DB	1211	CCGATCCGACCAAAATTCAGCTCATATAGACCCGACAGACTATTTTCAGTCTCTTAACAT	1270
Seq	1263	catlaccctaaatgcatlactgaagaagatlcacga	1268
DB	1271	CATTGCCCAAAATTCATATTCCTCCAGAACACTTATTCAGTCTCTTAACAT	1308
RESULT	3		
LOCUS	AK022580	2383 bp	mRNA
DEFINITION	Homo sapiens cDNA FLJ12518 f1s, clone NT2RM2001805.		PR1
ACCESSION	AK022580		
VERSION	AK022580.1	GI:10434050	
KEYWORDS	oligo coupling, f1s (full insert sequence),		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota: Mollusca: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.		
REFERENCE	1 (cites)		
AUTHORS	1sodai,T., Oda,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Saito,S., Matsuda,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Makamatsu,A., Nakamura,Y., Nagahori,I.K., Masuko,Y. and Sasaki,N.		
TITLE	NEBO human cDNA sequencing project		
JOURNAL	Unpublished (2000)		
REFERENCE	2 (bases 1 to 2383)		
AUTHORS	1sodai,T. and Otsuki,T.		
TITLE	Direct Submission		
COMMENT	Submitted (23-AUG-2000) to the DDBJ/EMBL/Genbank databases. Yako 1sodai, Helix Research Institute, Genomics Laboratory, 1532-3 Yano, Kisaragi, Chiba 292-0812, Japan (E-mail:genomics@uri.co.jp, Tel:81-438-3951, Fax:81-438-52-3952) NEBO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan. cDNA full insert sequencing: Research Association for biotechnology; cDNA library construction, 5'-63'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.		
FEATURES	Location/Qualifiers		
SOURCE	1..2383		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/cell_line="NT2"		
	/cell_type="testicular cDNA"		
	/clone="NT2RM2001805"		
	/clone_lib="NT2RM2"		
	/note="Cloning vector: pME18SP13-mRNA from uninduced NT2 neuronal precursor cells."		
	439..2343		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="BAH14110.1"		
	/db_xref="gi:10434051"		
	/translation="MSLVPTNNYLYTTLNOLKGGTVAVGVGVKPKRPYLSKGGDVC SVATVDTQTNKTLCLPSKGTVEALPTTKKDDVVRPIKALQVKKETQGTITSSGDP SLTEEGTAGVPIPTSSSKYTNFTEDIRKMYALIKVMSIMSPSWTLKLCIVPMK VFDLIDHILGKRAVYKASFLIAKVMGKPTTPPSWIKVLAQIVLEEDLSIHHLQNT D1LVYDNIHVAIRKLVKGSFPLSLIKRIQSMNSNCTPMISLEPHHIGTSYKGL VLPSPNSVDQDKDKIESANLVANDSHIVICSEEDPSFSGSVSLIEVERCOOLSL TLLDHOYLEPTVLCALIKOKAIDQYRIKALRSYKIRLPSQSVSLIEPKCGLDQEVY HEGLDDIFPKAKATPDKVAKQNTSKNSKNSVIVPKSGHDLLELDLSAPFLVGGTIIHKGCCSS CLLIIEGGLTSEICLSKRNKNSVIVPKSGHDLLELDLSAPFLVGGTIIHKGCCSS LKSLQNLNSLVKTSMTIPSVSAELGVIVLVGVVMTTLDLGGTGLAAVLMDSKPKP LQIPSVLEMDDDLOKSVDMIMDMCPICPKIDAVFWLRCFLKSNVNTGTNQLCYOY PFTVAEEV1"		
BASE COUNT	750 a	456 c	459 g
ORIGIN			718 t
Query Match			
	97.79%	Score:10416	DP:6; Length:2487;

[illegible]

AK001935 2077 bp mRNA PRI 22-FEB-2000
 LOCUS Homo sapiens cDNA J11073 f1s, clone PLACE1005026.
 ACCESSION AK001935
 VERSION AK001935.1 GI:7023508
 KEYWORDS oligo capping: f1s (full insert sequence).
 SOURCE Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE1
 clone:PLACE1005026.
 ORGANISM
 Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (sites)
 Isoqai,T., Ota,T., Hayashi,K., Suifu,Yama,T., Otsuki,T., Suzuki,Y.,
 Nishikawa,T., Naito,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,
 Matsumura,H., Ishii,S., Saito,K., Yamamoto,J.,
 Wakamatsu,A., Nakamura,Y., Nishihara,K., Masuho,Y. and Sasaki,N.
 NEDO human cDNA sequencing project
 unpublished (2000)
 2 (bases 1 to 2077)
 Isoqai,T. and Otsuki,T.
 Direct Submission
 Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases, Takao
 Isoqai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana,
 Kisarazu, Chiba 292-0812, Japan (E-mail:genomics-hri.co.jp,
 Tel:81-438-52-3951, Fax:81-438-52-3952)
 NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; cDNA full insert
 sequencing; Research Association for Biotechnology; cDNA library
 construction; 5'- & 3'- and one pass sequencing and clone selection;
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.

FEATURES

Location/Qualifiers
 1..2077
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="PLACE1005026"
 /clone_lib="PLACE1"
 /tissue_type="placenta"
 /note="Cloning vector: pME18SFLJ"
 131..2035
 /note="unmated protein product"
 /codon_start=1
 /protein_id="BA091948.1"
 /db_xref="GI:7023509"
 /translation="MSLVPAATNY IYTPDLNKGDT IVNVYGVVKKFFKPPYLSKTDVY:
 SVVT IVDQTNVKKLCLESGNYEALPT IYKNGD IYRHRK IYVYKRETCG IYSSGPA
 SIYFETGIGAP IYPTSSKYVNFTEHKKVVEALRYWASTHMSPSWTLIKLADVQPMQ
 YFDITCOILKKAEDVGSFLIKVMDKTRTPSPSWRVLIQDILVGLDLSH IURLNLIT
 DILVVDNHHVHARSLKVGSPILRYSLHTKIQSMNSNOTMLSLFFHLGGTSYGRGIR
 VLPSNSDVOLKDDLESANITANQHSVDICQSEPDSPSPSGSVSLYEVETQQLSA
 TLTLDHQLETPGICALIKOKATQOYIRAKRYSYKPRRLFQSVKYLHCPCGHLQVY
 HEGDDLI IYQIGATKTPDVKLQNTSLYDSK IWTTKNGRGRVAVHFKNG IYLPQNSH
 GLLI IYEGTLESEICLKNKENSIVIPVRSGRHDIJELLSAPFLIQCT IYHYGCKQSSS
 LRSIONLSHWDKTS IYPSVVAELG IYVLOYVYVMTFLDNGTGVLYEAY IYDSDKFF
 QIPASVYIMDDILQKSDVIMIMDPCIPG IYK IYDAYVPLECF IYKSNVNTGCTDNO IYGYI
 IYFTVAEDVI"
 657..401 c 193 a 626 b.

BASE COUNT

ORIGIN
 1079
 1080
 1081
 1082
 1083
 1084
 1085
 1086
 1087
 1088
 1089
 1090
 1091
 1092
 1093
 1094
 1095
 1096
 1097
 1098
 1099
 1100
 1101
 1102
 1103
 1104
 1105
 1106
 1107
 1108
 1109
 1110
 1111
 1112
 1113
 1114
 1115
 1116
 1117
 1118
 1119
 1120
 1121
 1122
 1123
 1124
 1125
 1126
 1127
 1128
 1129
 1130
 1131
 1132
 1133
 1134
 1135
 1136
 1137
 1138
 1139
 1140
 1141
 1142
 1143
 1144
 1145
 1146
 1147
 1148
 1149
 1150
 1151
 1152
 1153
 1154
 1155
 1156
 1157
 1158
 1159
 1160
 1161
 1162
 1163
 1164
 1165
 1166
 1167
 1168
 1169
 1170
 1171
 1172
 1173
 1174
 1175
 1176
 1177
 1178
 1179
 1180
 1181
 1182
 1183
 1184
 1185
 1186
 1187
 1188
 1189
 1190
 1191
 1192
 1193
 1194
 1195
 1196
 1197
 1198
 1199
 1200
 1201
 1202
 1203
 1204
 1205
 1206
 1207
 1208
 1209
 1210
 1211
 1212
 1213
 1214
 1215
 1216
 1217
 1218
 1219
 1220
 1221
 1222
 1223
 1224
 1225
 1226
 1227
 1228
 1229
 1230
 1231
 1232
 1233
 1234
 1235
 1236
 1237
 1238
 1239
 1240
 1241
 1242
 1243
 1244
 1245
 1246
 1247
 1248
 1249
 1250
 1251
 1252
 1253
 1254
 1255
 1256
 1257
 1258
 1259
 1260
 1261
 1262
 1263
 1264
 1265
 1266
 1267
 1268
 1269
 1270
 1271
 1272
 1273
 1274
 1275
 1276
 1277
 1278
 1279
 1280
 1281
 1282
 1283
 1284
 1285
 1286
 1287
 1288
 1289
 1290
 1291
 1292
 1293
 1294
 1295
 1296
 1297
 1298
 1299
 1300
 1301
 1302
 1303
 1304
 1305
 1306
 1307
 1308
 1309
 1310
 1311
 1312
 1313
 1314
 1315
 1316
 1317
 1318
 1319
 1320
 1321
 1322
 1323
 1324
 1325
 1326
 1327
 1328
 1329
 1330
 1331
 1332
 1333
 1334
 1335
 1336
 1337
 1338
 1339
 1340
 1341
 1342
 1343
 1344
 1345
 1346
 1347
 1348
 1349
 1350
 1351
 1352
 1353
 1354
 1355
 1356
 1357
 1358
 1359
 1360
 1361
 1362
 1363
 1364
 1365
 1366
 1367
 1368
 1369
 1370
 1371
 1372
 1373
 1374
 1375
 1376
 1377
 1378
 1379
 1380
 1381
 1382
 1383
 1384
 1385
 1386
 1387
 1388
 1389
 1390
 1391
 1392
 1393
 1394
 1395
 1396
 1397
 1398
 1399
 1400
 1401
 1402
 1403
 1404
 1405
 1406
 1407
 1408
 1409
 1410
 1411
 1412
 1413
 1414
 1415
 1416
 1417
 1418
 1419
 1420
 1421
 1422
 1423
 1424
 1425
 1426
 1427
 1428
 1429
 1430
 1431
 1432
 1433
 1434
 1435
 1436
 1437
 1438
 1439
 1440
 1441
 1442
 1443
 1444
 1445
 1446
 1447
 1448
 1449
 1450
 1451
 1452
 1453
 1454
 1455
 1456
 1457
 1458
 1459
 1460
 1461
 1462
 1463
 1464
 1465
 1466
 1467
 1468
 1469
 1470
 1471
 1472
 1473
 1474
 1475
 1476
 1477
 1478
 1479
 1480
 1481
 1482
 1483
 1484
 1485
 1486
 1487
 1488
 1489
 1490
 1491
 1492
 1493
 1494
 1495
 1496
 1497
 1498
 1499
 1500
 1501
 1502
 1503
 1504
 1505
 1506
 1507
 1508
 1509
 1510
 1511
 1512
 1513
 1514
 1515
 1516
 1517
 1518
 1519
 1520
 1521
 1522
 1523
 1524
 1525
 1526
 1527
 1528
 1529
 1530
 1531
 1532
 1533
 1534
 1535
 1536
 1537
 1538
 1539
 1540
 1541
 1542
 1543
 1544
 1545
 1546
 1547
 1548
 1549
 1550
 1551
 1552
 1553
 1554
 1555
 1556
 1557
 1558
 1559
 1560
 1561
 1562
 1563
 1564
 1565
 1566
 1567
 1568
 1569
 1570
 1571
 1572
 1573
 1574
 1575
 1576
 1577
 1578
 1579
 1580
 1581
 1582
 1583
 1584
 1585
 1586
 1587
 1588
 1589
 1590
 1591
 1592
 1593
 1594
 1595
 1596
 1597
 1598
 1599
 1600
 1601
 1602
 1603
 1604
 1605
 1606
 1607
 1608
 1609
 1610
 1611
 1612
 1613
 1614
 1615
 1616
 1617
 1618
 1619
 1620
 1621
 1622
 1623
 1624
 1625
 1626
 1627
 1628
 1629
 1630
 1631
 1632
 1633
 1634
 1635
 1636
 1637
 1638
 1639
 1640
 1641
 1642
 1643
 1644
 1645
 1646
 1647
 1648
 1649
 1650
 1651
 1652
 1653
 1654
 1655
 1656
 1657
 1658
 1659
 1660
 1661
 1662
 1663
 1664
 1665
 1666
 1667
 1668
 1669
 1670
 1671
 1672
 1673
 1674
 1675
 1676
 1677
 1678
 1679
 1680
 1681
 1682
 1683
 1684
 1685
 1686
 1687
 1688
 1689
 1690
 1691
 1692
 1693
 1694
 1695
 1696
 1697
 1698
 1699
 1700
 1701
 1702
 1703
 1704
 1705
 1706
 1707
 1708
 1709
 1710
 1711
 1712
 1713
 1714
 1715
 1716
 1717
 1718
 1719
 1720
 1721
 1722
 1723
 1724
 1725
 1726
 1727
 1728
 1729
 1730
 1731
 1732
 1733
 1734
 1735
 1736
 1737
 1738
 1739
 1740
 1741
 1742
 1743
 1744
 1745
 1746
 1747
 1748
 1749
 1750
 1751
 1752
 1753
 1754
 1755
 1756
 1757
 1758
 1759
 1760
 1761
 1762
 1763
 1764
 1765
 1766
 1767
 1768
 1769
 1770
 1771
 1772
 1773
 1774
 1775
 1776
 1777
 1778
 1779
 1780
 1781
 1782
 1783
 1784
 1785
 1786
 1787
 1788
 1789
 1790
 1791
 1792
 1793
 1794
 1795
 1796
 1797
 1798
 1799
 1800
 1801
 1802
 1803
 1804
 1805
 1806
 1807
 1808
 1809
 1810
 1811
 1812
 1813
 1814
 1815
 1816
 1817
 1818
 1819
 1820
 1821
 1822
 1823
 1824
 1825
 1826
 1827
 1828
 1829
 1830
 1831
 1832
 1833
 1834
 1835
 1836
 1837
 1838
 1839
 1840
 1841
 1842
 1843
 1844
 1845
 1846
 1847
 1848
 1849
 1850
 1851
 1852
 1853
 1854
 1855
 1856
 1857
 1858
 1859
 1860
 1861
 1862
 1863
 1864
 1865
 1866
 1867
 1868
 1869
 1870
 1871
 1872
 1873
 1874
 1875
 1876
 1877
 1878
 1879
 1880
 1881
 1882
 1883
 1884
 1885
 1886
 1887
 1888
 1889
 1890
 1891
 1892
 1893
 1894
 1895
 1896
 1897
 1898
 1899
 1900
 1901
 1902
 1903
 1904
 1905
 1906
 1907
 1908
 1909
 1910
 1911
 1912
 1913
 1914
 1915
 1916
 1917
 1918
 1919
 1920
 1921
 1922
 1923
 1924
 1925
 1926
 1927
 1928
 1929
 1930
 1931
 1932
 1933
 1934
 1935
 1936
 1937
 1938
 1939
 1940
 1941
 1942
 1943
 1944
 1945
 1946
 1947
 1948
 1949
 1950
 1951
 1952
 1953
 1954
 1955
 1956
 1957
 1958
 1959
 1960
 1961
 1962
 1963
 1964
 1965
 1966
 1967
 1968
 1969
 1970
 1971
 1972
 1973
 1974
 1975
 1976
 1977
 1978
 1979
 1980
 1981
 1982
 1983
 1984
 1985
 1986
 1987
 1988
 1989
 1990
 1991
 1992
 1993
 1994
 1995
 1996
 1997
 1998
 1999
 2000
 2001
 2002
 2003
 2004
 2005
 2006
 2007
 2008
 2009
 2010
 2011
 2012
 2013
 2014
 2015
 2016
 2017
 2018
 2019
 2020
 2021
 2022
 2023
 2024
 2025
 2026
 2027
 2028
 2029
 2030
 2031
 2032
 2033
 2034
 2035
 2036
 2037
 2038
 2039
 2040
 2041
 2042
 2043
 2044
 2045
 2046
 2047
 2048
 2049
 2050
 2051
 2052
 2053
 2054
 2055
 2056
 2057
 2058
 2059
 2060
 2061
 2062
 2063
 2064
 2065
 2066
 2067
 2068
 2069
 2070
 2071
 2072
 2073
 2074
 2075
 2076
 2077
 2078
 2079
 2080
 2081
 2082
 2083
 2084
 2085
 2086
 2087
 2088
 2089
 2090
 2091
 2092
 2093
 2094
 2095
 2096
 2097
 2098
 2099
 2100
 2101
 2102
 2103
 2104
 2105
 2106
 2107
 2108
 2109
 2110
 2111
 2112
 2113
 2114
 2115
 2116
 2117
 2118
 2119
 2120
 2121
 2122
 2123
 2124
 2125
 2126
 2127
 2128
 2129
 2130
 2131
 2132
 2133
 2134
 2135
 2136
 2137
 2138
 2139
 2140
 2141
 2142
 2143
 2144
 2145
 2146
 2147
 2148
 2149
 2150
 2151
 2152
 2153
 2154
 2155
 2156
 2157
 2158
 2159
 2160
 2161
 2162
 2163
 2164
 2165
 2166
 2167
 2168
 2169
 2170
 2171
 2172
 2173
 2174
 2175
 2176
 2177
 2178
 2179
 2180
 2181
 2182
 2183
 2184
 2185
 2186
 2187
 2188
 2189
 2190
 2191
 2192
 2193
 2194
 2195
 2196
 2197
 2198
 2199
 2200
 2201
 2202
 2203
 2204
 2205
 2206
 2207
 2208
 2209
 2210
 2211
 2212
 2213
 2214
 2215
 2216
 2217
 2218
 2219
 2220
 2221
 2222
 2223
 2224
 2225
 2226
 2227
 2228
 2229
 2230
 2231
 2232
 2233
 2234
 2235
 2236
 2237
 2238
 2239
 2240
 2241
 2242
 2243
 2244
 2245
 2246
 2247
 2248
 2249
 2250
 2251
 2252
 2253
 2254
 2255
 2256
 2257
 2258
 2259
 2260
 2261
 2262
 2263
 2264
 2265
 2266
 2267
 2268
 2269
 2270
 2271
 2272
 2273
 2274
 2275
 2276
 2277
 2278
 2279
 22

OS Oxytricha trifallax.
 OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Stichotrichida;
 CC Oxytrichidae; Oxytricha.
 OX NCBI_TaxID=5946;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Prescott J.D., Dubois M.L., Prescott D.M.;
 RT "Oxytricha trifallax micronuclear alpha telomere binding protein
 gene."
 RL Submitted (May-1998) to the EMBL/Genbank/DBJ databases.
 BK EMBL: AF067831; AAC27615.1;
 BK HSSB: P29549; 107C.
 BK InterPro: IPR003415; Telo_bind_alpha.
 BK Pfam: PF02307; Telo_bind_alpha.1.
 SQ SEQUENCE 497 AA; 56361 MW; 132C76BD6C3AH66 CRC64;
 Query Match 8.1%; Score 144; DB 5; Length 497;
 Best Local Similarity 21.9%; Pred. No. 0.00019;
 Matches 82; Conservative 57; Mismatches 142; Indels 94; Gaps 17;
 QY 7 TNYATPLN-QLKGGTVVNYGVVKKFPYVLSKCTVCGSVTVV-----Q 52
 DB 34 SKVEVVELTKQGLSVNQHVVAVVLDATPPYKTNQEHYICSLKIVDSLYLKKKCTGD 93
 QY 53 TNAVLTCLLSCNTHVADLTYKNGDLYVPHDKLQVYKKEITQG-----TSSGFASLTP-- 106
 DB 94 NSDVATVLYAKRFEDLPD1IHLCDD1HCHKATIRLNGCHQFNANIDYSSSMALFSTDK 153
 QY 107 -----KCTGAPD1PRT-SSKYFNTTEDIKRWALRVNASTIM-----SLWS 148
 DB 154 KSAIDIEIGGQVAT--SDLVPSHSCKNFTPEKRNAGIVQNIKKMAQGVPTGYVSINMF 211
 QY 149 TLKLCIDVQPMQYDPLTCGLAKAFVGCASPLAKVMKCTKTPFYSKVLTDHVLKSDLS 208
 DB 212 VALNFAVAGKQKQ-ENAVAKHGLHLEDEYIMKIKQSDAVI YTLALFKPT- 263
 QY 209 HNH-----RLONLTLD-----TLVYDNHVAKSLKVGSEFLKYSILTKLOSMSFNO 256
 DB 204 HLKCEEVKRIKSAITYDEISTOKKVLLSHYSNIYVFTISTSLAR--ELKGNITDDRAVER 321
 QY 257 TMLSLFPHL-----HCGTSYGCGLVNLTP-----SNSDVQLKKLLESANLT 298
 DB 322 AALKQDVSLSAVVLTEVDKKNAG-----LVNHSIDLPFHNDSDP--KELSKKPTF 369
 QY 299 ANQISDVICQSEPPD 313
 DB 370 KIQF--VYTKLEPSD 382
 RESULT 8
 Q45595 PRELIMINARY: PRT; 251 AA.
 AC 045595;
 DT 01-JUN-1998 (TREMBL) 06, Created)
 DT 01-JUN-1998 (TREMBL) 06, Last sequence update)
 DT 01-JUN-1999 (TREMBL) 09, Last annotation update)
 DE F57C2.3 PROTEIN.
 GN F57C2.3.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoida;
 CC Rhabditidae; Pseudocercaria; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Haynes C.;
 RL Submitted (Nov-1996) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-94150718; PubMed-7906398;
 RA Wilson R., Alnsough K., Anderson K., Haynes C., Herks M.,
 RA Bonfield J., Hurlon J., Connell M., Gopsey T., Cooper J., Toulson A.,
 RA Craxton M., Dear S., Du G., Durbin K., Fawcett A., Fulton L.,

RA Gardner A., Green P., Hawkins J., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirslen J., Laister R., Latreille P.,
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Ralton L., Roopar A., Saunders D., Shomkhen K.,
 RA Smardon N., Smith A., Sonhammer E., Staden K., Sultston J.,
 RA Thierrey-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterson K.,
 RA Watson A., Weinstock L., Wilkinson-Sprout J., Wohldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome 111 of C.
 elegans."
 RL Nature 368:32-38(1994).
 BK EMBL: Z83110; CAB05525.1;
 BK SEQUENCE 251 AA; 26987 MW; BP2P231D01C3BH9F CRC64;
 Query Match 7.4%; Score 140; DB 5; Length 251;
 Best Local Similarity 25.7%; Pred. No. 0.0013;
 Matches 52; Conservative 40; Mismatches 60; Indels 54; Gaps 10;
 QY 160 QYDPLTCGLAKAFVGCASPLAKVMK-----TTPPTPSKRV----- 196
 DB 25 RYFVLAQVHSVETNGQMLKVMGRGKCAEAERERKLPVTDSPKRYVTPNP 84
 QY 197 LDIVLVECDLSIDHMLONTLDLTVYDNHVAKSLKVGSEFLKYSILTKLOSMSFNO 256
 DB 85 RICKALFENC-----KELILETVYDHRIGLKNLNSGDVVALQVNVIA-----STRQ 132
 QY 257 TMLSLFPHL-GCTSYGKCTVQLTP--SNSDVQLKKLLESAN-----NLTANQISV 305
 DB 133 TEGV--LHCGASVQKQITTVYVDFEHPQNPKEVAVLTVAVDENTPEQGPPE 189
 QY 306 ICOS---EP-DPSFNGSLR 325
 DB 190 VALNVDHEDVQHEALKGVNLR 211
 RESULT 9
 Q9C2KB PRELIMINARY: PRT; 759 AA.
 ID Q9C2KB;
 AC Q9C2KB;
 DT 01-JUN-2001 (TREMBL) 17, Created)
 DT 01-JUN-2001 (TREMBL) 17, Last sequence update)
 DT 01-JUN-2001 (TREMBL) 17, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN 3H10.40.
 GN 3H10.40.
 OS Neurospora crassa.
 CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 CC Sordariaceae; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Alqn V., Holmset J., Brandt P., Farlman B., Holland K.,
 RA Nyakatura G., Mewes H.W., Mhammedi G.;
 RL Submitted (Feb-2001) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (Feb-2001) to the EMBL/Genbank/DBJ databases.
 DE EMBL: AL514442; CAC2864.1;
 SQ SEQUENCE 759 AA; 85218 MW; DC737A5674FP4F73 CRC64;
 Query Match 6.2%; Score 111; DB 3; Length 759;
 Best Local Similarity 20.7%; Pred. No. 0.29;
 Matches 63; Conservative 50; Mismatches 125; Indels 66; Gaps 12;
 QY 3 LVPATNYIYPTINOL-----KCTIVVNYGVVKKFPYVLSKCTVCGSVTVV-----V 51
 DB 18 LPTST---LTLKALIDDKQAGVSMNVICVAKDCAPVATHSDIKKCTHTLSDSLIF 74
 QY 52 QTNVKTCLLSCNTHVADLTYKNGDLYVPHDKLQVYKKEITQG1TSSGFASLFEETIG 111
 DB 75 DSSACVIAVLI-EPDAPNPVVAQVNVVLSAKVQKRNKRNKLSK1TTCVGVYKAAV 132

```

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK013364; BAB28810.1; -.
DR MGD; MGI:1920086; 2810458H16Rik.
SQ SEQUENCE 113 AA; 12740 MW; 00BBB9FC5D0C7088 CRC64;

Query Match      18.2%; Score 324; DB 11; Length 113;
Best Local Similarity 73.3%; Pred. No. 2.7e-21;
Matches 63; Conservative 6; Mismatches 17; Indels 0; Gaps 0;

QY 1 MSIVPATNYIYTPLNQLKGGTIVNVYGVVKKFKPPYLSKGTGYCSVVTIVDQTNVKTCL 60
DB 1 MSAPVAPSTYTPLELLKGGTVNVYGVVKKFKPPYLSKGTGYCSVVTIVDQTNVKTCL 60

QY 61 LFSGVEALPIIYKNCIDVRFHRLKI 86
DB 61 LFSRDYDLPVYKVGDIIVCFQCLAV 86

RESULT 5
OL3988
ID OL3988 PRELIMINARY; PRT: 555 AA.
DT 01-JAN-1999 (TREMBlrel. 09, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE HYPOTHETICAL 64.1 KDA PROTEIN C26H5.06 IN CHROMOSOME 1.
GN SPAC26H5.06.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycotaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972.
RA Oliver K., Harris D., Bartell B.G., Rajandream M.A., Wood V.;
RL Submitted (Sep-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z99126; CAB16192.2; -.
KW Hypothetical protein.
SQ SEQUENCE 555 AA; 64111 MW; A79DAA95A0C4F803 CRC64;

Query Match      9.4%; Score 167.5; DB 3; Length 555;
Best Local Similarity 24.4%; Pred. No. 1.8e-06;
Matches 77; Conservative 54; Mismatches 118; Indels 67; Gaps 15;

QY 18 KCGTIVNVYGVVKKFKPPYLS-KCT-DYCSVVTIVDQT-----NVKLTCLSGNVEALPI 71
DB 38 KNTIVNLGIVGIVDTPSQSLGRTKDWVTIVYLMDPTDTSISGIGIILHLSKGNLIPV 97

QY 72 IYKNGDIVRFHRLKIQVYKKTGGTSSGFA-----SLTFEGTLCGAPIIPRTS----- 119
DB 98 IKQVGOPLLLHUIITLRSYRDRTQGLSKDQFRYALWPKDFSSNSKDTLCPOPPRLMKTKGDK 157

QY 120 -----SKYFNFTEDIKHVEALRWVAS-----THMS-PSWTLKLCDDVQPMQYFDLTC 166
DB 158 EEQFALLLNKINDEQTNKIKNGELLSSARONQTLGSPSVFSLLSQITPHQRCSEFYA 217

QY 167 QLLGKAEVDGASFLKLVNDGTR-----TPPS---WRVLIQDLVLEGLDLSIHRQLNL 216
DB 218 QVI-KTWYSDKNFTLYVDYTENELFFPMSYTSRRWR-----GPPG-----RF 261

```

```

QY 217 TIDILVYDNHIVHARS-LKVCSEFLRITYSLHTKLOSMHSENOTMLSEFLHIGGTS--YGR 273
DB 262 SIRCILWDEHDFYCRNTIKGGDYVVMKNVTKIDILG-----YLCILHGDSAKRYNM 314

QY 274 GIRVLPESDSVDQLK 289
DB 315 STEKVDSEEPENEIK 130

RESULT 6
Q9FNH7
ID Q9FNH7 PRELIMINARY; PRT: 463 AA.
AC Q9FNH7;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE GB|AAD29059.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98069011; PubMed=9405947;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tsubata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,052 bp covered by thirteen
RT physically assigned pl clones.";
RL DNA Res. 4:291-300(1997).
DR EMBL; AB006700; BAB08953.1; -.
SQ SEQUENCE 463 AA; 53630 MW; A735DBF109155D7D CRC64;

Query Match      8.6%; Score 153; DB 10; Length 463;
Best Local Similarity 23.7%; Pred. No. 2.8e-05;
Matches 79; Conservative 49; Mismatches 138; Indels 68; Gaps 13;

QY 23 VNVYGVVKKFKPPYLSKGTGYCSVVTIVD--QTNVKTCLLIFSNYEALPIYKNGDIVR 80
DB 25 VNLGIVIVELG---FSNGDCSCETLKIVDWYSGSGLPVKFVART(RDLP)RVESIGDIL 81

QY 81 FHLKIQVYKKTGGL-----TSSGFASLTFEGTLCGAPIIPRTSKYFNFTEDIKHVEAL 116
DB 82 LSRVKIVLHNRKITALCNETTSSSFA--LNGKHISVDS(IPYSS)PRFLMHFDKNFLSNL 139

QY 137 RVWASTIM---SPSWTLKLCDDVQPMQYFDLTCGLGKAEVDGASFLKLVNDGTRTPP 192
DB 140 REMMITTYKFEEDGGCGCTSLK-DLEKGECSNLSCQIVHLSKVYKDRWYLFVNDGTE--MP 195

QY 193 SWRVLIQD---LVLEGLDLSIHRQLNLITDILVYDNHIVHARS-LKVCSEFLRITYSLHTKL 248
DB 196 PCNTLVRKSERLPCVDEPLTPYNGRKP-----TGSVLR-----I 233

QY 249 QSMHSENOTMLSEFLHIGGTSYGRGIRVLPESNSDSVDQLKDLIESANITANQHSIDVQC 308
DB 234 VDRVSEKQAIFHCIO-----PCQHVKLL-----NLFFQVNMGLWNATFTTSTKMOYMS 281

QY 309 SEPDDSPNGVSLRIP-----PCWSSVAR 331
DB 282 REMEAFSPQRFLLTPFLILCMGKPKSPRWNPILAR 315

RESULT 7
Q76380
ID Q76380 PRELIMINARY; PRT: 497 AA.
AC Q76380;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ALPHA TELOMERE BINDING PROTEIN.

```


FT DOMAIN 188 567 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 244 546 PROTEIN KINASE.
 FT BIND 250 258 ATP (BY SIMILARITY).
 FT BINDING 277 277 ATP (BY SIMILARITY).
 FT ACT_SITE 379 379 BY SIMILARITY.
 FT CARBOHYD 70 70 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 388 389 KN -> RS (IN REF. 2).
 FT CONFLICT 403 403 R -> G (IN REF. 2).
 FT CONFLICT 405 405 D -> S (IN REF. 2).
 FT CONFLICT 477 477 K -> R (IN REF. 2).
 SQ SEQUENCE 567 AA: 64241 MW: 6117642A51A3B75 CRC64; EC1D7642A51A3B75 CRC64;

Query Match 5.0%; Score 89.5; DB 1; Length 567;
 Best Local Similarity 20.7%; Pred. No. 4.8;
 Matches 69; Conservative 51; Mismatches 136; Indels 77; Gaps 15;

QY 33 KPPYLSKGTGYGVTVTDQTNVKTCL-LFSGNYEALPI----IYKNGDIVRHLKIQ 87
 DB 205 KPKLMDSNCALILEDSDISSTCANNININTELLPIELDTLVCKG---RF-----AE 257
 QY 88 VYKKTGCTSGFASLTFTGTLAPTIP-----RTSSKYPN-----FTTE 128
 DB 258 VYKAKLKQNTSQE-----TVAVKIPPEEYSSMKTEKDIFSDINLKHENILQELTAE 311
 QY 129 DHKWEALRVWASTIMSPWTLKLCQVQPMOYEDLTCLLCKAEVDGASELLKXWDGTR 188
 DB 312 ERKTEMGQYWLITAFHAKGNLQYELTRHIVISWEDL---RKLGSLANGLAHLHS--DHTP 367
 QY 189 TFPFSRVLIODL-----VLKCDLS-----HRLONLTIDILVYDVHIVHVARSLKV 235
 DB 368 CGRPKMPIVHRDKSSNVLKNDLTCCLDGCLSLKLDPTLSVDDLSANGSGVGTARYMAP 427
 QY 236 GSFLRLYSHTLKQSMNSNOT-MSLEFHLIGHTS-----YGRGTRVLPE 280
 DB 428 ---EVLESRMNLENMESFKQTDVYSMALVLEWMTSRGNVAGEVKDYEPFGSKVREHPC 483
 QY 281 SMSVDQDLKKDESANLTA-----NOHSDVLCOS 309
 DB 484 VESKDNVLRDKRPEIPDSFWLNHOGTQVCTET 516

RESULT 13
 TYPHL MYCGE STANDARD; PRT; 421 AA.
 AC P47297: Q49312;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THYMIDINE PHOSPHORYLASE (EC 2.4.2.4) (TDRPASE).
 GN D60A OR MG051.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID:20477;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 43530 / G-37;
 RX MEDLINE=96026346; PubMed=7569993;
 RA Fraser C.M., Gockayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bull C.L., Korfman A.R., Sutton G., Kelley J.M.,
 RA Fleischman J.L., Weidman J.F., Smell K.V., Sandusky M., Fuhrmann J.L.,
 RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Bertick J.M.,
 RA Tomb J.-F., Dougherty D.A., Bott K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RT "The minimal genes complement of Mycoplasma genitalium";
 RL Science 270:397-403(1995).
 RN 121
 RP SEQUENCE OF 388-421 FROM N.A.
 RC STRAIN-ATCC 43530 / G-37;
 RX MEDLINE=94075230; PubMed=8253680;
 RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;

RT "A survey of the Mycoplasma genitalium genome by using random sequencing."; J. Bacteriol. 175:7918-7930(1993).
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -1- FUNCTION: THE ENZYMES WHICH CATALYZE THE REVERSIBLE PHOSPHORYLOSIS OF PYRIMIDINE NUCLEOSIDES ARE INVOLVED IN THE DEGRADATION OF THESE COMPOUNDS AND IN THEIR UTILIZATION AS CARBON AND ENERGY SOURCES, OR IN THE RESCUE OF PYRIMIDINE BASES FOR NUCLEOTIDE SYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: THYMIDINE + PHOSPHATE -> THYMINE + 2-DEOXY-D-RIBOSE 1-PHOSPHATE.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: BELONGS TO THE THYMIDINE/PYRIMIDINE-NUCLEOSIDE PHOSPHORYLASES FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U39684; AAC71267.1;
 DR EMBL: U02191; AAD12476.1;
 DR TIGR: MG051;
 DR InterPro: IPR003262; Anthr-phosphoribosyltransf.
 DR InterPro: IPR000312; Glycos_transf_3.
 DR InterPro: IPR000053; Thymid_phosphils.
 DR Pfam: PF00591; Glycos_transf_3; 1.
 DR ProDom: PD001864; Glycos_transf_3; 1.
 DR ProDom: PD005916; Thymid_phosphils; 1.
 DR PROSITE: PS00647; THYMID_PHOSPHORYLASE; 1.
 KW Transferase; Glycosyltransferase; Complete proteome.
 FT CONFLICT 392 392 V -> I (IN REF. 2).
 SQ SEQUENCE 421 AA: 46355 MW: 649CD517CD6E3E62 CRC64;

Query Match 5.0%; Score 99; DB 1; Length 421;
 Best Local Similarity 22.1%; Pred. No. 3.5;
 Matches 68; Conservative 48; Mismatches 114; Indels 78; Gaps 15;

QY 14 LNLKCGCTIVNYGVGVKFKPIYLSKGTGYGVTVTD-OTNVKLTCLLFSG-NYEALPI 71
 DB 7 LNKQRKALNL-AETNWFVNAVLANK-----TIADYQITAFILMAIFKGMNPELPL 57
 QY 72 LYK-----NGDIVRFP-HRLKIQVKKPTQGI-----TSSGFASLTFTGTLGAP 114
 DB 58 LTKAMVDTCETLKFNIHCKLSVDKISTGIGIKVSLALVPIITSLGFS----- 105
 QY 115 IPTSSKYENFTTEDIKMVEALRVWASTIMSPWTLKLCQVQPMOYEDLT-COLLGK-- 171
 DB 106 VAKLSGRGLGYTGCTIDKLEAVGV-----KTELTDQQAQACLDKNDCTF 154
 QY 172 --AEVDCASFLKVMKTRTPPEPSMKVLFQDLVLEKDLSHHRLONLTIDILVYDVHIV 229
 DB 155 DIAPVDKVLVGLRDTFTGTVDSLP-----GLASSIMSKKLLA-----VMNEY 197
 QY 230 ARSLKVGSGFLRYSLHTKL-----QSMNSRNOTMLSLFELHICGTSGYGRGTRVLPESNSIV 285
 DB 198 DLKYGKGAFCRTKKLANELAKLMQSLAKSFKKLSVKLS-DMNQVLGKAV;NVLVNEAV 256
 QY 286 DQILKKDLE 293
 DB 257 NFIKQDL 264

RESULT 14
 NUSM PARTE STANDARD; PRT; 570 AA.
 ID NUSM PARTE
 AC P15584;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).

FT	DOMAIN	360	403	LDL-RECEPTOR CLASS A 4.	FT	DISULFID	1295	1304	BY SIMILARITY.
FT	DOMAIN	404	504	IG-LIKE C2-TYPE DOMAIN 1.	FT	DISULFID	1307	1322	BY SIMILARITY.
FT	DOMAIN	521	530	LAMININ EGF-LIKE 1 (N-TERMINAL).	FT	DISULFID	1563	1572	BY SIMILARITY.
FT	DOMAIN	531	730	LAMININ DOMAIN IV 1 (DOMAIN III A).	FT	DISULFID	1565	1579	BY SIMILARITY.
FT	DOMAIN	731	763	LAMININ EGF-LIKE 1 (C-TERMINAL).	FT	DISULFID	1582	1591	BY SIMILARITY.
FT	DOMAIN	764	813	LAMININ EGF-LIKE 2.	FT	DISULFID	1594	1610	BY SIMILARITY.
FT	DOMAIN	814	871	LAMININ EGF-LIKE 3.	FT	DISULFID	1613	1628	BY SIMILARITY.
FT	DOMAIN	879	923	LAMININ EGF-LIKE 4 (INCOMPLETE).	FT	DISULFID	1615	1638	BY SIMILARITY.
FT	DOMAIN	924	933	LAMININ EGF-LIKE 5 (N-TERMINAL).	FT	DISULFID	1641	1650	BY SIMILARITY.
FT	DOMAIN	934	1125	LAMININ DOMAIN IV 2 (DOMAIN III B).	FT	DISULFID	1653	1668	BY SIMILARITY.
FT	DOMAIN	1126	1158	LAMININ EGF-LIKE 5 (C-TERMINAL).	FT	DISULFID	1792	1839	BY SIMILARITY.
FT	DOMAIN	1159	1208	LAMININ EGF-LIKE 6.	FT	DISULFID	1886	1932	BY SIMILARITY.
FT	DOMAIN	1209	1265	LAMININ EGF-LIKE 7.	FT	DISULFID	1976	2021	BY SIMILARITY.
FT	DOMAIN	1275	1324	LAMININ EGF-LIKE 8.	FT	DISULFID	2073	2118	BY SIMILARITY.
FT	DOMAIN	1325	1334	LAMININ EGF-LIKE 9 (N-TERMINAL).	FT	DISULFID	2170	2215	BY SIMILARITY.
FT	DOMAIN	1335	1529	LAMININ DOMAIN IV 3 (DOMAIN III C).	FT	DISULFID	2268	2313	BY SIMILARITY.
FT	DOMAIN	1530	1562	LAMININ EGF-LIKE 9 (C-TERMINAL).	FT	DISULFID	2365	2413	BY SIMILARITY.
FT	DOMAIN	1563	1612	LAMININ EGF-LIKE 10.	FT	DISULFID	2456	2506	BY SIMILARITY.
FT	DOMAIN	1613	1670	LAMININ EGF-LIKE 11.	FT	DISULFID	2554	2599	BY SIMILARITY.
FT	DOMAIN	1677	1771	IG-LIKE C2-TYPE DOMAIN 2.	FT	DISULFID	2641	2686	BY SIMILARITY.
FT	DOMAIN	1772	1865	IG-LIKE C2-TYPE DOMAIN 3.	FT	DISULFID	2831	2876	BY SIMILARITY.
FT	DOMAIN	1866	1954	IG-LIKE C2-TYPE DOMAIN 4.	FT	DISULFID	2917	2962	BY SIMILARITY.
FT	DOMAIN	1955	2049	IG-LIKE C2-TYPE DOMAIN 5.					
FT	DOMAIN	2050	2148	IG-LIKE C2-TYPE DOMAIN 6.					
FT	DOMAIN	2149	2244	IG-LIKE C2-TYPE DOMAIN 7.					
FT	DOMAIN	2245	2343	IG-LIKE C2-TYPE DOMAIN 8.					
FT	DOMAIN	2344	2436	IG-LIKE C2-TYPE DOMAIN 9.					
FT	DOMAIN	2437	2532	IG-LIKE C2-TYPE DOMAIN 10.					
FT	DOMAIN	2533	2619	IG-LIKE C2-TYPE DOMAIN 11.					
FT	DOMAIN	2620	2720	IG-LIKE C2-TYPE DOMAIN 12.					
FT	DOMAIN	2721	2809	IG-LIKE C2-TYPE DOMAIN 13.					
FT	DOMAIN	2810	2895	IG-LIKE C2-TYPE DOMAIN 14.					
FT	DOMAIN	2896	2980	IG-LIKE C2-TYPE DOMAIN 15.					
FT	DOMAIN	2981	3130	LAMININ G-LIKE 1 (GLOBULAR DOMAIN V A).					
FT	DOMAIN	3049	3241	EGF-LIKE 1.					
FT	DOMAIN	3304	3495	EGF-LIKE 2.					
FT	DOMAIN	3558	3705	LAMININ G-LIKE 2 (GLOBULAR DOMAIN V B).					
FT	SITE	65	67	HEPARAN SULFATE (POTENTIAL).					
FT	SITE	71	73	HEPARAN SULFATE (POTENTIAL).					
FT	SITE	76	78	HEPARAN SULFATE (POTENTIAL).					
FT	SITE	3615	3617	MEDIATES MOTOR NEURON ATTACHMENT (POTENTIAL).					
FT	DISULFID	199	212	BY SIMILARITY.					
FT	DISULFID	206	225	BY SIMILARITY.					
FT	DISULFID	219	234	BY SIMILARITY.					
FT	DISULFID	285	297	BY SIMILARITY.					
FT	DISULFID	292	310	BY SIMILARITY.					
FT	DISULFID	304	319	BY SIMILARITY.					
FT	DISULFID	325	337	BY SIMILARITY.					
FT	DISULFID	332	350	BY SIMILARITY.					
FT	DISULFID	344	359	BY SIMILARITY.					
FT	DISULFID	368	381	BY SIMILARITY.					
FT	DISULFID	375	394	BY SIMILARITY.					
FT	DISULFID	388	403	BY SIMILARITY.					
FT	DISULFID	428	470	BY SIMILARITY.					
FT	DISULFID	764	773	BY SIMILARITY.					
FT	DISULFID	766	780	BY SIMILARITY.					
FT	DISULFID	783	792	BY SIMILARITY.					
FT	DISULFID	795	811	BY SIMILARITY.					
FT	DISULFID	814	829	BY SIMILARITY.					
FT	DISULFID	816	839	BY SIMILARITY.					
FT	DISULFID	842	851	BY SIMILARITY.					
FT	DISULFID	854	869	BY SIMILARITY.					
FT	DISULFID	1159	1168	BY SIMILARITY.					
FT	DISULFID	1161	1175	BY SIMILARITY.					
FT	DISULFID	1178	1187	BY SIMILARITY.					
FT	DISULFID	1190	1206	BY SIMILARITY.					
FT	DISULFID	1209	1224	BY SIMILARITY.					
FT	DISULFID	1211	1234	BY SIMILARITY.					
FT	DISULFID	1237	1246	BY SIMILARITY.					
FT	DISULFID	1249	1263	BY SIMILARITY.					
FT	DISULFID	1275	1287	BY SIMILARITY.					
FT	DISULFID	1277	1293	BY SIMILARITY.					

Query Match

5.2%; Score 92.5; DB 1; Length 3707;

Best Local Similarity 19.1%; Pred. No. 38;

Matches 72; Conservative 56; Mismatches 141; Indels 107; Gaps 15;

QY	49	IVDOTNVKLTCLFSGNYEALPIIYKNGDIVRFH-----RLKIQVYKKETOG-----I 96
DB	2063	VAEQGTLDLNCVV-PGHAHAQVTHMKRGGSGLPTHGTHGSLRLYQVSSADSGEYVGSVL 2121
QY	97	TSSG-----FASLTFGTGLCAPTRTSKYENFTFEDHKMVEALRYWASTH 143
DB	2122	SSSGPLASVLVSTPAANVHIPGVVPPRIETSSSRVAEGOTLDLSCV----VPCQAH 2177
QY	144	MSPSW-----TLKLCQVQPMQYFDLTCGLLGAENYDCAFLAAVWATR 198
DB	2178	AQVTHMKRGGSGLPAGHOVHGHLNRSVSPADSGFYCVTCSGTTLEASVLT 2437
QY	189	TPPFSWRVLTDLVLEGLDSIIHHLNLTIDLVYDN-HVHIV-----ARSLKVG 237
DB	2238	SP1PA-PGLAQPVY[ESSSSHLTEGQTVDLKCVVPQAHQVTHMKRGSSI,PAHQHTIGS 2296
QY	238	FLRLYSLHTK-----LOSMMSENQTMLSLEFHLGGTSYGRGIRVLIPESNSVDQL 288
DB	2297	LRLYQLSPADSGEYVGVAGSSHPEHSEPKL-----TVPSSONSSFR 2341
QY	289	KK----DLESAMITANQHSOV-----ICSEPPDSSP---PNC--VSLR 322
DB	2342	RSPVISIEPPSPSTVOODASPKCLIHGEGAMPKVEWKRIRDOLEEDRHVHISPRGSIITIV 2401
QY	323	PGCWSSVARSRLAAS 338
DB	2402	APGPATWPTACVASN 2417

RESULT 11

ALU1_HUMAN

ID ALU1_HUMAN STANDARD; PRT; 591 AA.

AC P39188;

DT 01-FEB-1995 (Rel. 31, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DE ALU SUBFAMILY J SEQUENCE CONTAMINATION WARNING ENTRY.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95021758; PubMed=7935834;

RA Claverie J.-M., Mukalowski W.;

RT "Alu alert.";

FT	D0M4IN	414.5	417.8		EEF-LIKE 4.
PT	SUTPE	429.3	433.1		LAMININ G-LIKE 3 (POTENTIAL).
FV	SITE	6.5	6.7		HEPARAN SULFATE (POTENTIAL).
FV	SITE	7.1	7.3		HEPARAN SULFATE (POTENTIAL).
FV	SITE	7.6	7.8		HEPARAN SULFATE (POTENTIAL).
FT	SITE	41.51	41.53		MEDIANES MOTOR NEURON ATTACHMENT (POTENTIAL).
FT	SITE	430.1	430.3		MEDIANES MOTOR NEURON ATTACHMENT (POTENTIAL).
FT	D1S0L.F1D	199	212		HY SIMILARITY.
FT	D1S0L.F1D	206	225		HY SIMILARITY.
FT	D1S0L.F1D	219	244		HY SIMILARITY.
FT	D1S0L.F1D	285	297		HY SIMILARITY.
FT	D1S0L.F1D	292	310		HY SIMILARITY.
FT	D1S0L.F1D	404	419		HY SIMILARITY.
FT	D1S0L.F1D	325	337		HY SIMILARITY.
FT	D1S0L.F1D	342	350		HY SIMILARITY.
FT	D1S0L.F1D	344	359		HY SIMILARITY.
FT	D1S0L.F1D	368	381		HY SIMILARITY.
FT	D1S0L.F1D	375	394		HY SIMILARITY.
OY	Match	5, 8%	Score: 102.5%.	DB 1:	Length 4 494
Match	Seq. Similarity	24, 1%	Pred. No. 6-63		
Match	B4: Conserved	49%	Matched: 140;	Ind-Is	B5: Gaps
OY	4B TTVGCTNNATTCCLTSGNVEALDIYKRNQI-----VFPHIKLKJ-----GVKKKE	9.2			
DB	2257 TVAGCGDTLLSLGV--ACGMADGVWYKRKGCSLDARIHQVGSILYLVSASPLADGGVCMA	2.15			
OY	9.3 YGGLTSSGASLTFFRGTGA-----PI-IHRTSKRYNRFTTD-----	1.29			
DB	2316 SNGMEAS--ITVTVTGYGCANLAYTAGSTOPIKLEPFSSQVAEGCOTLLDLNCVVPQSIAQ	2.37			
OY	130 ---HKKNVALRWASTWMSPTMLTKICGVDPYQVDPYICOLIGKAKMGASPLTKVMIG	1.86			
DB	2374 VNMHRKGSLSYPKHIOHTGS---LIRQVANSVADSEYCYRGVAGSVPLEASVLVTIEFA	24.29			
OY	187 TWTP----PPSRRLVLIQDLLEGDSIIHLRNLTDILVYN-HVHV-----AR	2.31			
DB	2430 GSVPALKAVIPYVFI-----ESSSQVAECOTLLDLNCIVACDAIVWMHRKGSIDAR	24.82			
OY	234 SLKYSPFLKI-----YSILPRLOSMMNSINDMTLSFELILGGTSYGCGIIVLPES	2.81			
DB	2483 HGCVHSKRLLOVTVADSEYCYRGVAGSVCTDASVLYTLQQLSGSHSGCAVYVRIES	254.2			
OY	282 NSVDQLKKDLKSANIYAMQSDVIYCGSDPSFN-----GVSLRP	3.24			
DB	2543 SSASLANGHTLIDLNCIVASQAVHETTWYRKGSLSLRDQVCSRLRIIP	2590			
RESULT	8				
YH47_METJA	YH47_METJA	STANDARD:	PKT:	462 AA.	
AC	OS547:				
PT	01-NOV-1997 (Rel. 35, Created)				
PT	01-NOV-1997 (Rel. 35, Last sequence update)				
PT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	HYPOTHEICAL PROTEIN M01147.				
CN	M01147.				
G5	Melhanococcus jannaschii.				
OC	Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;				
CC	Methanococcus.				
OX	NCHI_TaxID:2190;				
NN	111				
KP	SEQUENCE FROM N.A.				
KP	STRAIN-JAL-1 / DSM 2661 / ATCC 43067;				
KX	MEDLINE=96337999; PubMed=868087;				
RX	Bult C.J., White O., Olsen G.J., Zhou B., Fleischmann K.D.,				
KA	Sutton G.C., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,				
KA	Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,				
KA	Overbeek R., Kirkness E.F., McIninch K.G., Merrick J.M., Glodok A.,				
KA	Scott J.D., Goodenough N.S.M., Weidman J.F., Fuhlmann J.D., Nguyen D.,				
KA	Hittorffack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hannon M.C.,				

```

or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: X01804; CAA25945.1;
CC DR EMBL: M15080; AAA32516.1;
CC DR EMBL: AF158101; AAD42472.1;
CC DR EMBL: M01160; AAC05392.1;
CC DR PIR: A04298; NCBPX6.
KW Hydrolase; Nuclease; Exonuclease; ATP-binding; DNA repair.
FT NP_BIND 36 43 ATP (POTENTIAL).
SQ SEQUENCE 560 AA: 63613 MW: D1078294.1 BAB2PG CIR64;

Query Match 5.8%; Score 102.5; DB 1; Length 560;
Best Local Similarity 21.0%; Pred. No. 0.36;
Matches 63; Conservative 46; Mismatches 98; Indels 93; Gaps 12;

UY 65 NYEALPIYKNGDIVREHRLKIQVYKKEGTIT-----SSGFASLTFFECTLGAPIPR 117
DB 3 NFKLRVYKYNIMSVQNGCIDIQDKVOKTLTGNNCGGKSTMLEAITF-GFGKPPRDV 61
UY 118 TSSKYFNFTEDUKWVEALRWVASTHSPSWTLKLCDDVPMQV-----F 162
DB 62 KKGCOLINSTNKKELVE---LW-----MEYDEKKYYLKRGQKNVF 99
UY 163 DLTC-----OLICKAEVYKASFLKLYWDGTR--TPE-----PSWRV 196
DB 100 EITVNGTRLNESASSKDFQAFEQIG---MSYASFQKIVVLGAGTYTPMGLSTPARRK 156
UY 197 LTQDLVLQEDLSHILRLNLCTIDLVYDNHIVHVARSLKVSFLRYSLSLTKLQ-SMNSEN 255
DB 157 LVEDLLEVTGLAEMDKLNKALIRELNSQNVLDVKKDSITQIKIYNDNVEROKKLTCDN 216
UY 256 QTMLSLEFLILGCTGYGRCIRVLPSNSVDVQLKKDLSEANLTANQHSIDVICQSEPDPSF 315
DB 217 UTRLO-----NNYDDLAKAEARTKSEITEAN---ERLVNIVLDEDDPTDAF 258

RESULT 7
PCBM_HUMAN
ID PGHM_HUMAN STANDARD: PRT: 4393 AA.
AC P98160; Q16287;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE
DE PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92112994; PubMed-1740768;
RA Kallunki P., Tryggvason K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor."
RL J. Cell Biol. 116:559-571(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-92235084; PubMed-1569102;
RA Murdoch A.D., Dodge G.R., Cohen I., Tian R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor."
RL J. Biol. Chem. 267:8544-8557(1992).
RN [3]
RP SEQUENCE OF 1018-1472 FROM N.A.
RX MEDLINE-92235084; PubMed-1569102;
RA Murdoch A.D., Dodge G.R., Cohen I., Tian R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor."
RL J. Biol. Chem. 267:8544-8557(1992).
RN [4]
RP SEQUENCE OF 892-1398 FROM N.A.
RX MEDLINE-92112994; PubMed-1740768;
RA Kallunki P., Tryggvason K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor."
RL J. Cell Biol. 116:559-571(1992).
RN [5]
RP SEQUENCE OF 1-21 FROM N.A.
RX MEDLINE-94052171; PubMed-8234307;
RA Cohen I.R., Graessle S., Murdoch A.D., Iozzo R.V.;
RT "Structural characterization of the complete human perlecan gene and
RT its promoter."
RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
CC +1- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
CC MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
CC CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
CC PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
CC SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
CC SUBSTRATE FOR CELLS.
CC -1- SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
CC DIMERS OR STELLATE STRUCTURES.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC -1- +3-4: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
CC -1- SIMILARITY: CONTAINS 22 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: X62515; CAA44373.1;
CC DR EMBL: M65289; AAA52700.1;
CC DR EMBL: M64283; AAA52699.1;
CC DR EMBL: S76436; AAU21121.2;
CC DR EMBL: L22078; NOT_ANNOTATED_CDS.
CC DR USSP: P00740; L1XA.
CC DR MIM: 142461;
CC DR InterPro: IPR000561; EGF-like.
CC DR InterPro: IPR000742; EGF_2.
CC DR InterPro: IPR001438; EGF_11.
CC DR InterPro: IPR003006; Ig_MHC.
CC DR InterPro: IPR003598; Ig_C2.
CC DR InterPro: IPR002172; LDL_recept_A.
CC DR InterPro: IPR000034; Laminin_B.
CC DR InterPro: IPR002049; Laminin_EGF.
CC DR InterPro: IPR001791; Laminin_G.
CC DR Pfam: PF000082; SEA.
CC DR Pfam: PF00008; EGF; 4.
CC DR Pfam: PF00047; Ig; 22.
CC DR Pfam: PF00052; Laminin_B; 3.
CC DR Pfam: PF00053; Laminin_EGF; 8.

```

RN [11]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-9221884; PubMed-1313446;
 RA Wang F.L., O'Farrell S., Clayberger C., Kinsley A.M.,
 RT "Identification and molecular cloning of lactiferin, a novel human T
 cell activation antigen that is a member of the 14 gene
 superfamily."?
 RL J. Immunol. 148:2600-2608(1992).
 CC -1- FUNCTION: MAY BE INVOLVED IN ADHESIVE INTERACTIONS OF ACTIVATED T
 AND NK CELLS DURING THE LATE PHASE OF THE IMMUNE RESPONSE. MAY
 FUNCTION AT A TIME AFTER T AND NK CELLS HAVE PENETRATED THE
 ENDOTHELIUM USING INTEGRINS AND SELECTINS, WHEN THEY ARE ACTIVELY
 ENGAGING DISEASED CELLS AND MOVING WITHIN AREAS OF INFLAMMATION.
 CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
 CC -1- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED ON NORMAL T CELL LINES AND CLONES,
 AND SOME TRANSFORMED T CELLS, BUT NO OTHER CHLONED CELL LINES.
 CC TESTED: IT IS EXPRESSED AT VERY LOW LEVELS ON ACTIVATED T CELLS.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT LOW LEVELS ON PERIPHERAL T CELLS
 AND IS STRONGLY UP-REGULATED AFTER ACTIVATION, PEAKING 6 TO 9 DAYS
 AFTER THE ACTIVATING STIMULUS.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC -1- DATABASE: NMR-PRO; NOT-CD guide C96b entry;
 CC WWW-URL: <http://www.ncbi.nlm.nih.gov/ncbi/cd/c96b.htm>.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
 or send an email to license@sib-sb.ch).
 CC -----
 DR EMBL: M86282; AAA36662.1;
 DR PIR: A46462; A46462.
 DR InterPro: IPR003599; 19.
 DR InterPro: IPR003006; 19_MHC.
 DR Pfam: PF00047; 19; 3.
 DR SMART: SM00409; 16; 1.
 RW Immunoglobulin domain; glycoprotein; Transmembrane; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 569 T-CELL SURFACE PROTEIN TACTILE.
 FT DOMAIN 22 501 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 502 525 POTENTIAL.
 FT DOMAIN 526 569 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 534 486 PRO/SER/THR-RICH.
 FT DOMAIN 538 547 PRO-RICH.
 FT DOMAIN 38 125 IG-LIKE V-TYPE DOMAIN 1.
 FT DOMAIN 156 238 IG-LIKE V-TYPE DOMAIN 2.
 FT DOMAIN 267 346 IG-LIKE C2-TYPE DOMAIN.
 FT DISULFID 45 118 PROBABLE.
 FT DISULFID 163 231 PROBABLE.
 FT DISULFID 274 339 PROBABLE.
 FT CARBOHYD 42 42 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 107 107 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 156 156 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 166 166 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 261 261 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 262 262 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 284 284 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 334 334 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 419 419 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 481 481 N-LINKED (GLCNAC. .) (POTENTIAL).
 SO SEQUENCE 569 AA; 63887 MW; DF2F3DE356F3BF2 CRC64;

Query Match

6.1k; Score 109; DB 1; Length 569;

Host Local Similarity 20.7%; Pred. No. 0.1;
 Matches 78; Conservative 71; Mismatches 145; Indels 82; Gaps 20;
 DB 16 QKACCTLVNYYGVKPKPKYLSKTHYC-----SVWTVV--GTNKKLFCIL----- 61
 DB 55 QMMSKRYNKLDELAVHYGV-----GYCAVGHGCHSLVTFEPENCSKMLHLKNSG 310
 QY 62 -FSGNYVALDLYKKNQDLYVPHQKLOVKKKTKGTPSSCSASTPRTGTCATLHKRSS 120
 DB 111 GVSQKTKHMLVTFHLYVTFVRLVLYVTVADW RSNITLLELRGTLCTVFNSSS 166
 QY 121 KY-----PNTPEHKKVVALVWASVTHMSVSTLAK-----LGVGPHGVDF-- 163
 DB 170 KISEPFAVMSVINGNQTETL ISQHLISNCTLDKRWKLCITVRLHLSVQLDQKOR 227
 QY 164 -LITGVL-IGAEVYVASTLAKVMTFRTPTSKRVLG---LVLEADP-----LSNTHL 213
 DB 228 KPSCHHVCCHNKLKSSPYVRAKLEP---VIVENSTVVALVKKPTVLAKEVPR 282
 QY 214 QNTE--LV-LVYVNDV-VIVARSLKVG--SEPLKYSILKLS--MNSNDQTMSLAT 263
 DB 283 ANTWFTMGSLADEKRGVITPEPKRGKMLKLSVLTFRHSNKPALSDMTITCM-- 440
 QY 264 ILHCCSYGHCIVLIDPSNSDVQIKRDLASANIYANQHSVDYCGSEPHDPSVNCVSLAD 323
 DB 341 -----ALSTVPCNKV-----WNLSSEKTPPLALGHSSTDPPLSVSTSLPTGP 384
 QY 324 PCMSVAKSLDLAST 339
 DB 385 SVASVSAVAVYVTS 400
 RESULT 6
 EX02_1474 STANDARD; PRT; 560 AA.
 AC P04552;
 DT 13-AUG-1987 (rel. 0%; Created)
 DT 01-JAN-1990 (rel. 13; Last sequence update)
 DT 20-AUG-2001 (rel. 40; Last annotation update)
 DE EXONUCLEASE SUBUNIT 2 (EC 3.1.11.1) (PROTEIN GP46).
 GN 46.
 OS Bacteriophage T4.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
 OC T4-like phages;
 OX NCBI_TaxID:10465;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-85257446; PubMed-4018026;
 RA Gram H., Kueger W.,
 RT "Genes 55, alpha 9L, 47 and 46 of bacteriophage T4: The genomic
 organization as deduced by sequence analysis."?
 RL EMBO J. 4:257-264(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kutter E., Arisaka F., Tsubota A., Mosig C.,
 RA Mesyanzhinov V., Kueger W., Stidham T., Thomas H.,
 RT "Bacteriophage T4 genome analysis."?
 RL Submitted (J01-2000) to the EMBL/Genbank/DDBJ databases.
 RN [3]
 RA Hsu T., Karam J.;
 RP SEQUENCE OF 478-560 FROM N.A.
 CC Submitted (APR-1987) to the EMBL/Genbank/DDBJ databases.
 CC -1- FUNCTION: EXONUCLEASE INVOLVED IN PHAGE DNA RECOMBINATION,
 REPLICATION, AND REPAIR.
 CC -1- SUBUNIT: CONSISTS OF TWO SUBUNITS: GP47 AND GP46.
 CC -1- SIMILARITY: STRONG TO T5 PROTEIN D13 AND TO YEAST RAD52.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>


```

A:Reference number: A86141; MUID:21016719
A:Accession: D96723
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1198 <STO-
A:Cross-references: CH:AK005173; MJD:g2194131; P1DN:AAH61106.1; GSTR:CN0014
C:Genetics:
A:Gene: P2015.20
A:Map position: 1

```

Query Match	5, 58;	Score	98;	DB	2;	Length	1198;
Best local Similarity	21, 48;	Pred.	No. 7;				
Matches	65;	Conservat	46;	Mismatch	119;	Indels	44;
							Gaps
							15;

QY	50	VDOINWALCTCLD-----SGN EALILYKNGD VRFHRLNOVKRQCTGTSNGASVAF	106
Db	769	MEECNLPICLFEPKDESGRGSVAVYVILGHIIDFQSGKSLQKMS-----LHSPALNLI	B22
QY	107	ICCTGCA--P-IHPPTSSKRYFNFTTEHDHKEV EALUVMVASTHSPISW--TLKLCYVOPMGYF	162
Db	823	EDELADDEPKRPVAKIYSTYVALAGSDMSWVLTETKRWQVYVYRTGCPPIALALCALPLR--	B79
QY	163	DITCGILLGRAEVIGASFLRLVMDICTKTPFSWIKVLIQDVLIECHSHIRIQNTIDILIV	222
Db	880	EIMENILBPKKDM-----QW-----DKLASEELAGSLWRKLT-----	909
QY	223	YDINIVIVARSLKXGSLFKLYSLITTK--LQSMNSFNQPMILSLFPIILHCTGTSYGCTIVLPES	281
Db	910	-----SLFTWS--WSLWLVVWVLSLSDHDEY-----LHMSHETDMDAAQYKQD	949
QY	282	NSIDVQLKRLK-----LHSAVLTANQHSIVICGSEIDPDSPPKAVSLHPPQMS	327
Db	950	RSKISRLKMSSECFGRYKKELEFANLTVEKMKRLKASR--PDLNRLDGLTDSQAK	1006
QY	328	SVAR	331
Db	1007	DFMR	1010

```

RESULT 13
T06149
hypothetical protein F2437.120 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #Sequence_revision: 30-Apr-1999 #Lexl_change: 15-Sep-2000
C:Accession: T06149
R:Bevan, M.; Vitale, D.; Liguori, R.; Arquier, A.; De Simone, V.; Bancroft, J.; Mewes
submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15493
A:Accession: T06149
A:Molecule type: DNA
A:Residues: 1-474 <BEV>
A:Cross-references: EMBL:AL021768; GSPDB:GN00062; ATSP: F2437.120
A:Experimental source: cultivar Columbia; BAC clone F2437
C:Genetics:
A:Gene: ATSP:F2437.120
A:Map position: 4
A:Introns: 73/3; 102/3; 144/3; 205/2; 253/1
A:Superfamily: Arabidopsis thaliana hypothetical protein F2437.161

```

[illegible]

QY	151	-----LKEICVPOI-----MOYEDLH-COHLGRKAVDCDAS?-----LLEV	163
DB	217	HHHAACA111AAELFTYVIDLOSTREYLCOEFDLTHCOLEHINWVVALSTAFSDJKAQOILEL	276
QY	184	MOGSTRKPPHS-----WVVLIDEN-----LESDLSH1HRLQNL77DLIVY	223
DB	277	YERK1PTSOEKSVESSGCAVAVHOF1SDRMASKTEKCSSDIEGSSQVMI.SQS-----	328
QY	224	DNI-VH-VAKSLKRVSPFLKIVSLATPKLQSNMSEKQTM1SLEPH1HGCTSYGKGRVLPES	291
DB	329	DHLSHMGKSRSLIC-----EYNSFEAKRMLDHSVVC-----N1WVER	467
QY	282	NSDVPD--QULKRDLE294	
DB	368	SDIVCVVQVQKRLQ481	

RESULT 14
 H64443
 Hypothetical protein M1147 - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence: revision 13-Sep-1996 #exch: 21-Jul-2000
 C:Accession: H64443
 R:Mail: G.-J. White, G. J. Olson, G.-J. Zhou, L.-J. Fritschmann, R.D. Sutton, G.G. Black
 J. Reich, C.L. Overberg, R. K. Krieger, E.F. McInerck, R.G. Merrick, J.M. Glodde,
 Tson, J.D. Sadow, P.W. Hanna, M.C. Colton, M.D. Roberts, K.M. Hurst, M.A.
 Saiter, 275, 1056-1073, 1996
 A:Authors: Kaine, B.P.; Koryotovsky, M.; Klink, B.P.; Frost, C.M.; Smith, H.O.; Woese
 A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*
 A:Reference number: Ab4100; PMID: 86343999
 A:Accession: H64443
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residue: 1-462 - bp.
 A:Cross-references: GI:467557; GI:477117; NID:4159177; P108:AA99156.1; P10:4159177
 C:Genetics:
 A:Map position: FOR1086043-1087791
 A:Start codon: TTT

Query Match	5, 37	Score	94.0%	Dh	27	Identical	462%
blast Local Simlaid	19, 88%	Prod. No.	3, 5%				
Matches	64	Conservative	58	Mismatches	120	Indels	81
						Gap	14
QY	16	QLKGGT-----	VNVGV-----	VKFF-RHY-----	LSKTDYGSVTTV	50	
Db	59	ELKGGLERKDEIVVNI	KVIDIYNVEFNCDFEFVFMEEYQIV	SVFNCNAYFVITF-	117		
QY	51	DDNNVKKIQLLRSCNVEAL	IIYK-----	NGDIVRHQKLOVKKFQGTSSG	100		
Db	118	KQSVTFRTINQVDVPI	DI FEENATVSTAKGNI	INFSG---TIPKE---	SH	167	
QY	101	FASLTVEPTLGAII	PIRTSSKYFNTEDHKVDA	RYMASSTHMSPTWTLK	LCQVPPQ	160	
Db	168	KSTPEEGMTVSVTPPI	NAEYNSFESHVYFDDI	-----	SFNLSPTDCKRQ	217	
QY	161	YFDLTGCLLKAEVACAS	FLKVMKCTKTPPYSKV	LQDL-----	VLEGDSIIHKLQ	214	
Db	218	-DYSEFKIDKENFKGAI	FLK-----	TQPLNHTTIIENQ	LSTKSPFLKTDVREV---	265	
QY	215	NLTIDIIIVVDNIVHAK	SLKVGSPILKITYLTKIK	OSNMSENQMLSL	EPHICGTGCGG	274	
Db	266	-LLCQVKKKEEII	LSHKIKIKESGAKDKDILN	KLKELGLSTYK	IIIDQFRKSVLAETYN	324	
QY	275	IVVLPESNSVDOLK	DLPSANI	297			
Db	325	LRISLENN-----	RYYEASNI	341			
RESULT	15						
0500349							
exonuclease' EC 3.1.1.1) sbcd - Escherichia coli							
Species: Escherichia coli							

Db 138 IKDEGDDFTYRSAYSGKYNFTDOKLLKNTRAWKSYFAKN-----DV----- 183
 QY 164 LTCOLLCKAEVDGASFLKAVWCDTRTPPSWRVLLODLVLEGLDLSHRLQLNLTIDILVY 223
 Db 184 -----IDEMVTPUSQARQEGDFNVGKVQIVRHRYDYYTSDLRVK 224
 QY 224 DN-----HVRVARSKVGSLRIY-SLHTKQSMNSNOTMLSLFHLHGTSYGRGIRV 277
 Db 225 DTSKATWFLTVSR-----KFPRLYEGVVIKIRSVNIDSE-----TERERCLLEL 268
 QY 278 LPESN-----SDVDLKKDLES 294
 Db 269 APHSNIMTFPFSRLAKSLDS 289

RESULT 5
 S22581
 telomere-binding protein alpha chain - Stylonychia mytilus
 C:Species: Stylonychia mytilus
 C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 07-Dec-1999
 C:Accession: S22581
 R:Fanig, G.; Cech, T.R.
 Nucleic Acids Res. 19, 5515-5518, 1991
 A:Title: Molecular cloning of a telomere-binding protein genes from Stylonychia mytilus.
 A:Reference number: S22580; MUID:92051290
 A:Accession: S22581
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-493 <FAN>
 A:Cross-references: EMBL:X61749; NID:q10194; PTD:q578501
 C:Genetics:
 A:Genetic code: SCC5
 A:Introns: 32/3
 C:Keywords: DNA binding; nucleus

Query Match 7.64; Score 135; DB 2; Length 493;
 Best Local Similarity 21.18; Pred. No. 0.0014;
 Matches 79; Conservative 57; Mismatches 148; Indels 90; Gaps 15;

QY 6 ATNYITPLNQ--LKGCT IVNVYGVVKKFPYLSKGTGYGVVTIVD----- 51
 Db 30 STRYKRVVNLKASLTSAEAQHFYGVVIDATPEYKTNQERYICSLKVVDPSPSLKSKQCTG 89
 QY 52 OTNVKLTCLLSCNVEALPIYKNGD IVRFIRLKLQVYKHTQ-----GTTSSG 100
 Db 90 DASDVTILVYAKRFEDLPIIIRIGDITVRHATIRLYNGQORQFNARVFNSSWALFSTD 149
 QY 101 FASLTFECLTGAIPRT-----SSKYENFTTEDHKMVEALRVWASTIUM-----SPSWT 149
 Db 150 KKSALQKICGGPASPDLTPPAFSCKNYTPFKSEANLIONIRKWAQVFOQYNYVSSDMPT 209
 QY 150 LKRLGIDVQPMOYFDLTCLLAKAEVDGASFLKAVWCDTRTPPSWRVLLODLVLEGLDLSH 209
 Db 210 PLNKAAQAKGD-FQVAKILQVFEFLDEYTNELKLDQSGQVF-----YTLALKLKFPPI 261
 QY 210 HI-----RIQNIITD-----ILVYDNIHVARSKVGSLRIYSLHTKQSMNSNOT 257
 Db 262 LRAGRVVLRATYDTSSTOKKVLALLSHYSNTVTFVSASKLAK-EVKAKVTIDKRVKKA 319
 QY 258 MSLSEFHL-----HGTSYGRGIRVLP-----SNSVDQLKKDLESANCTA 299
 Db 320 ALKQDVLSLAVVLEVDKKHAG-----LPTHSLODLFHNADTD---KELSSKUTFR 367
 QY 300 NOHSDVICQSEPD 313
 Db 368 TQF--YITRVEPD 379

RESULT 6
 T22839
 hypothetical protein F57C2.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T22839
 R:Baynes, C.
 submitted to the EMBL data library, November 1996
 A:Reference number: Z19624
 A:Accession: T22839
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-251 <WIL>
 A:Cross-references: EMBL:Z83110; PIDN:CAB05525.1; GSPDB:UN00020; CESP:F57C2.3
 A:Experimental source: clone F57C2
 C:Genetics:
 A:Gene: CESP:F57C2.3
 A:Map position: 2
 A:Introns: 115/3; 198/2

Query Match 7.34; Score 130; DB 2; Length 251;
 Best Local Similarity 25.74; Pred. No. 0.0014;
 Matches 52; Conservative 10; Mismatches 66; Indels 54; Gaps 10;

QY 160 QYEDLTCLLCKAEVDGASFLKAVWCD-----TKTFPPSMRV----- 196
 Db 25 RYEDLVLAQVHVVETTINGAQLRWGRKFGPGASAEKRERRLFIHVTDQSFKRYIVPPNP 84
 QY 197 LIODLVLEGLDLSHRLQLNLTIDILVYDNIHVARSKVGSLRIYSLHTKQSMNSNOT 256
 Db 85 RICKATEENG-----KELLIDVDYDHRGLKWLNSGDFVATONVHAA-----STRQ 132
 QY 257 TMLSEFHLH-GGTSYGRGIRVLP-----SNSVDQLKKDLES-----NLTANOIISDV 305
 Db 133 TEMQV---LIGGASVQGTITTPVDPFEHAEQNFKKKKEAVLETVAIDENFTFOQPEE 189
 QY 306 ICOS-----EP-DDSPFNGVSLR 322
 Db 190 VAENHVDEEPQEEALPRGVVLR 211

RESULT 7
 A16462
 T cell activation antigen Tacile - human
 C:Species: Homo sapiens (man)
 C:Date: 18-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
 C:Accession: A16462
 R:Wang, P.L.; O'Farrell, S.; Clayberger, C.; Krensky, A.M.
 J. Immunol. 148, 2600-2608, 1992
 A:Title: Identification and molecular cloning of Tacile. A novel human T cell acti-
 A:Reference number: A16462; MUID:92218864
 A:Accession: A16462
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-569 -WAN.
 A:Cross-references: GB:MBR282; NID:q18671; PIDN:AAA36662.1; PTD:q18672
 A:Note: sequence extracted from NCHI backbone (NCBIIN:94434, NCBIIP:94435)
 C:Keywords: Transmembrane protein

Query Match 6.14; Score 109; DB 2; Length 569;
 Best Local Similarity 20.74; Pred. No. 0.28;
 Matches 78; Conservative 71; Mismatches 145; Indels 85; Gaps 20;

QY 16 OLKGGTIVNVYGVVKKFPYLSKGTGYC-----SVVTIVD--OTNVKLTCLL----- 61
 Db 55 QMWSKVTKNIDLIAYVHPOV---GFCYACRPECISLVTTETTFENGSKWTLHRLNMSC 110
 QY 62 FSGRYEALPIYKNGD IVRFIRLKLQVYKHTQITSSGFASLTFECLTGAIPRTSS 120
 Db 111 SVSGRYECMLVLYPEIOTKIYVNLITQHTADEW-NSNHTLEIENOTLEIPCFQNSSS 169
 QY 121 KY-----ENFTEDHKMVEALRVWASTIUMSPSWTLK-----LCDVQPMQVFD---- 163
 Db 170 KISSEFTYAWSVEDNCTOETL--ISQNILNSTLLKDRVKLGTDYRLHLSFVQIFDDGR 227

S35524
 telomere-binding protein, 51K - Euplotis crassus
 C:Species: Euplotis crassus
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-Dec-1999
 C:Accession: S35524
 R:Wang, W.; Skopp, R.; Scofield, M.; Price, C.
 Nucleic Acids Res. 20, 6621-6629, 1992
 A:Title: Euplotis crassus has genes encoding telomere-binding proteins and telomere-binding reference number: S35524; MUID:93126105
 A:Reference number: S35524
 A:Accession: S35524
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-446 <MAN>
 A:Cross-references: EMBL:M96819; NID:q159023; P1D:q159024
 C:Genetics:
 A:Genetic code: SGC9
 A:Introns: 9/3
 C:Keywords: DNA binding; nucleus

Query Match 8.4%; Score 167; Dh 2; Length 446;
 Best Local Similarity 21.3%; Prev. No. 2,30-06;
 Matches 91; Conservative 65; Mismatches 142; Indels 140; Gaps 19;

8 NYLYPLNOLK--GGTVNNGVGVKFKPYLSKGTG--YGSVVTIVGQTNV-----K 56
 11 HYQYSLSTKKKGEHGYHFCVVIDASPPV--KCEKRVVGVKVDHSSVAKGCKLNT 68
 57 LKCLFSGVNEALPIYKNGDIYRPHRLQVYKKEFGCIVSSGPA--LTFEGTIGAP 113
 69 VNVFFSOFEDLPJLQVGDIVRHAKALQYNIAMKQINVMNYVSSMCLTICNDKELP 128
 114 IIPRT-----SKYF-----NFTTEIDH--KKEVALKQWASTHMSWTL--LKIC 154
 127 GGGVYNNENGTNNYSYTYTYPNSKSPYQGHCTKLKDKKWSKDYNSNNVAVQVKA 188
 155 DVQPMQY---FDLTGCLGKAEVDCASFLKVMQCT-----KTPSPWVLIQDL 201
 189 DLETAMKNKTDPLAKVETISDNDQYNTVGSINSTGCTGTGHLFKKFTPI----- 240
 202 VLKGLSHLHRLQNLITDILVYDNIHVAKSLKVGSPFLKYSFLKLSQMSSENOITMLST 261
 241 LKQGVLRKYSVAKEDNSLIFSSHSNLIK-----FESFSIHKLS-----SI 285
 262 EPHLGGTSGYGCIVLPESSNDVQOLK-----DLSANIT----- 298
 286 SSDTHIKTCVTK-IDKAAHMKMDITPLKLPFNPKSKLPSQPSVLKVDYTKNLEDYVG 344
 299 -----ANOHSVJGSEPPD-SFNGVSLRPPG 325
 345 AFDGKKHYSCKKTPKDAELRWNLKILVTQYKNGQDDKAVMIHLDNSFTKGIN--VAN 402
 326 MSSVAKSR 333
 11: 1 :
 Db 403 WSNATTK 410

RESULT 3
 A41221
 macronuclear alpha protein (alanine version) - Oxytricha nova
 C:Species: Oxytricha nova
 C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 02-Mar-2001
 C:Accession: A41221
 R:Gray, J.T.; Celander, D.W.; Price, C.M.; Cech, T.R.
 Cell 67, 807-814, 1991
 A:Title: Cloning and expression of genes for the Oxytricha telomere-binding protein: spcA; reference number: A41221; MUID:92035001
 A:Accession: A41221
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-495 <GHA>
 A:Cross-references: GH:M68931; NID:q159811; P1D:q159812
 A:Accession: B41221

A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-310 'S', 412-455, 'H', 457-495, <GNC>
 A:Cross-references: GH:M68930; NID:q159813; P1D:q159814
 C:Genetics:
 A:Genetic code: SGC9
 A:Introns: 18/3; 377/3
 C:Keywords: DNA binding; nucleus

Query Match 8.2%; Score 146.5; Dh 2; Length 495;
 Best Local Similarity 23.5%; Prev. No. 0,00015;
 Matches 80; Conservative 52; Mismatches 133; Indels 75; Gaps 14;

26 YGVVKEFKRPYLSKGTGCSVVTIVGQTNV-----NVRKTLGLFSGNVEALPIY 74
 55 VAVVIDATPTTYKTNQPKRTIGSLIVDPHTYLKQCKGACDASIVATIVYAKRPHEDPIH 114
 74 KNGDIYRPHRLKQVYKKEFGT-----GTFSSGPASTPEPTIGAP11PRT----- 118
 115 KNGDIYRPHRLKQVYKKEFGT-----GTFSSGPASTPEPTIGAP11PRT----- 118
 119 SKRYTPTTEIDHRLKQVYKKEFGT-----GTFSSGPASTPEPTIGAP11PRT----- 118
 175 SKRYTPTTEIDHRLKQVYKKEFGT-----GTFSSGPASTPEPTIGAP11PRT----- 118
 172 AFDGKASFLKVMQCTKTPPTSMWTLIDPLVLCGLSHH-----RLQNLITD----- 219
 244 HDLDPTNEHLKIDASGGV-----YTLSLKLETHVIRGCVVRLKSVTYDPTSTQK 286
 220 ILV---YNIHIVAVASIKVGSFLKYSFLKLSQMSSENOITMLSTEPHLGGTSGYGCIV 276
 287 VLLSHYSNLIPTIQSLSLAKELR-----AKIGDHSEVAVSLKRNVLNA-----V 343
 277 VLPSSNIDVQOLKELGSSANI--TANQHSVJGSEPPDSE 315
 334 VLTE-----VKEKHAQDPSTSLQDLIHADSKHLQAGVTF 369

RESULT 4
 S35525
 telomere-binding protein - Euplotis crassus
 C:Species: Euplotis crassus
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-Dec-1999
 C:Accession: S35525
 R:Wang, W.; Skopp, R.; Scofield, M.; Price, C.
 Nucleic Acids Res. 20, 6621-6629, 1992
 A:Title: Euplotis crassus has genes encoding telomere-binding proteins and telomere-binding reference number: S35524; MUID:93126105
 A:Accession: S35525
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-460 <MAN>
 A:Cross-references: EMBL:M96819; NID:q159025; P1D:q159026
 C:Genetics:
 A:Genetic code: SGC9
 A:Introns: 18/3; 377/3
 C:Keywords: DNA binding; nucleus

Query Match 8.0%; Score 142.5; Dh 2; Length 460;
 Best Local Similarity 22.7%; Prev. No. 0,00029;
 Matches 73; Conservative 45; Mismatches 118; Indels 85; Gaps 13;

9 YLYPLNOL--KCGTVNNGVGVKFKPYLSKGTGCSVVTIVGQTNV-----LTC 59
 19 YEYTEIGSIEENFASINFAVAVVDMCFPYKVDKRYKMCYLKVIDTTHVKKGGDNFAIV 76
 60 LKCLFSGVNEALPIYKNGDIYRPHRLKQVYKKEFGT-----SGTFASLTTEGTIGAP1 114
 79 ALQSKRFEEDLPJLQVGDIVRHAKALQYNIAMKQINVMNYVSSMCLFSADEVAPEV 137
 115 IP-----RTSKRYFNTTTEIDHRLKQVYKKEFGT-----V 163

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:09:30 ; Search time 38.59 seconds
(without alignments)
671.141 Million cell updates/sec

Title: US-09-816-248-15
Perfect score: 1776
Sequence: 1 MSLVPATNVIYTPNLQKGG.....LRPPCWSSVARSRLIAASTS 340

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	167.5	9.4	535	2 T38425	hypothetical prote
2	167	9.4	446	2 S35524	telomere-binding p
3	146.5	8.2	495	2 A41221	nucleonuclear alpha
4	142.5	8.0	460	2 S35525	telomere-binding p
5	135	7.6	493	2 S22581	telomere-binding p
6	130	7.3	251	2 T22839	hypothetical prote
7	109	6.1	569	2 A46462	T cell activation
8	109	6.1	587	2 T22006	hypothetical prote
9	103	5.8	1044	2 T00342	hypothetical prote
10	102.5	5.8	560	1 NCBPX6	exonuclease 46 (EC
11	102.5	5.8	4391	2 A38096	perlecan precursor
12	98	5.5	1198	2 D96723	hypothetical prote
13	95.5	5.4	474	2 T06149	hypothetical prote
14	94.5	5.3	462	2 B64443	hypothetical prote
15	93.5	5.3	400	1 JS0349	exonuclease (EC 3.
16	93.5	5.3	400	2 D85535	ATP-dependent dsbN
17	92.5	5.2	3707	2 S19252	heparan sulfate pr
18	90.5	5.1	610	2 G84306	biotin carboxylase
19	90.5	5.1	1030	2 F96763	hypothetical prote
20	90	5.1	361	2 E86692	N-acetylglucosyl-L
21	90	5.1	937	2 B86210	protein F2G5.6 [l
22	89.5	5.0	528	2 T32012	hypothetical prote
23	89.5	5.0	567	2 JN0459	transforming growt
24	89	5.0	421	1 F64205	thymidine phosphor
25	89	5.0	570	2 S07744	NADH dehydrogenase
26	89	5.0	683	2 B71325	conserved hypothet
27	88.5	5.0	1217	2 T21403	hypothetical prote
28	88	5.0	3788	2 T30851	lysosomal traffick
29	88	5.0	3796	2 T18514	lysosomal traffick

30	87.5	4.9	553	2 T52362	hypothetical prote
31	87	4.9	4196	2 T41274	dynein heavy chain
32	86.5	4.9	318	2 A84466	hypothetical prote
33	86.5	4.9	592	2 S51371	transforming growt
34	86.5	4.9	1002	2 T46033	receptor protein k
35	86.5	4.9	1428	1 L5BYT2	DNA topoisomerase
36	86	4.8	899	2 C84765	hypothetical prote
37	85.5	4.8	574	2 A96612	hypothetical prote
38	85.5	4.8	861	2 T15903	protein kinase C h
39	85	4.8	426	1 N71249	hypothetical prote
40	85	4.8	2285	2 T12796	probable transglyc
41	84.5	4.8	334	2 F70384	hypothetical prote
42	84.5	4.8	507	2 S15476	transmembrane glyc
43	84.5	4.8	749	2 H82691	topoisomerase IV s
44	84.5	4.8	905	1 RCBYS5	regulatory protein
45	84.5	4.8	1886	2 S04921	nuclear pore prote

ALIGNMENTS

RESULT 1

T38425
hypothetical protein SPAC26H5.06 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T38425
R:Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z21748
A:Accession: T38425
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-535 <ALL>
A:Cross-references: EMBL:Z99126; P1DN:GAB16192.1; GSPDB:GN00086; SPL:SPAC26H5.06
A:Experimental source: strain 972h-; cosmid c26H5
C:Genetics:
A:Gene: SPDB:SPAC26H5.06
A:Map position: 1
A:Introns: 44/1
C:Superfamily: Schizosaccharomyces hypothetical protein SPAC26H5.06

Query Match	9.4%	Score 167.5;	DB 2;	Length 535;
Best Local Similarity	24.4%	Pred. No. 2.7e-06;		
Matches	77;	Conservative	54;	Mismatches 118;
				Indels 67;
				Gaps 15;
QY	18	KGGTIVNVYGVVKKPKPPYLS-KGT-DYGSVVTIVDQT-----NVKLTGCLLPSGNVEALP1	71	
DB	18	KKNTIVNLPFGIVKDETPSRQSLHGTCKDWVTTVYLMDPTGDTSSICGQIHLPKQCGNDLPV	77	
QY	72	LYKNGDIVRFHRLKIQVKKETQGITSSGFA-----SLTEGTGAPLIPRTS-----	119	
DB	78	IKOVQPLALHQLTKRSYKDRTOGLSKDQFRYALMDPFSNKSNDKDFLACDPMPLMKTGDK	137	
QY	120	-----SKYFNPTTEDIKMFVZALRVWAS-----THMS-PSMTLLKICADVQPMQVDFLFC	166	
DB	138	EEQFALLINKIWEQTNKIKNGELLSTSSARONGTGLSVPSVSFSLISQITTHQRCSFYA	197	
QY	167	QLLGAKEVDGASFLKLVMDTR-----TPPFS---WRVLQDILVEGDLSHLHLQML	216	
DB	198	QVL-KTWSDKNFTLVYVDTYENELFPMSPYTSRNR-----CPFG-----RF	241	
QY	217	TIDILVONHIVVARS-LKVGSLRLYLSLITKLOSNSENQTMLSLEFLHKGTS--YGR	273	
DB	242	SLRCLMWDGHDYCYRNYIKEGDYVVKVNRTRKIDHLG-----YLECFLHGDSSAKRYNM	294	
QY	274	GIRVLPSNSDVOQLK	289	
DB	295	STEKVDSEEPNELIK	310	
RESULT	2			

```

PK 27-JUL-1999; 990S-0145914.
PK 28-JUL-1999; 990S-0145951.
PK 02-AUG-1999; 990S-0146386.
PK 02-AUG-1999; 990S-0146386.
PK 02-AUG-1999; 990S-0146386.
PK 03-AUG-1999; 990S-0147034.
PK 04-AUG-1999; 990S-0147204.
PK 04-AUG-1999; 990S-0147302.
PK 05-AUG-1999; 990S-0147260.
PK 06-AUG-1999; 990S-0147303.
PK 06-AUG-1999; 990S-0147416.
PK 09-AUG-1999; 990S-0147493.
PK 09-AUG-1999; 990S-0147943.
PK 10-AUG-1999; 990S-0148171.
PK 11-AUG-1999; 990S-0148319.
PK 12-AUG-1999; 990S-0148341.
PK 13-AUG-1999; 990S-0148564.
PK 16-AUG-1999; 990S-0149368.
PK 17-AUG-1999; 990S-0149175.
PK 18-AUG-1999; 990S-0149426.
PK 20-AUG-1999; 990S-0149722.
PK 20-AUG-1999; 990S-0149723.
PK 20-AUG-1999; 990S-0149929.
PK 23-AUG-1999; 990S-0149902.
PK 23-AUG-1999; 990S-0149930.
PK 25-AUG-1999; 990S-0150566.
PK 26-AUG-1999; 990S-0150864.
PK 27-AUG-1999; 990S-0151065.
PK 27-AUG-1999; 990S-0151066.
PK 27-AUG-1999; 990S-0151080.
PK 30-AUG-1999; 990S-0151303.
PK 31-AUG-1999; 990S-0151438.
PK 01-SEP-1999; 990S-0151930.
PK 07-SEP-1999; 990S-0152363.
PK 10-SEP-1999; 990S-0153070.
PK 13-SEP-1999; 990S-0153758.
PK 15-SEP-1999; 990S-0154018.
PK 16-SEP-1999; 990S-0154039.
PK 20-SEP-1999; 990S-0154779.
PK 22-SEP-1999; 990S-0155139.
PK 23-SEP-1999; 990S-0155486.
PK 24-SEP-1999; 990S-0155659.
PK 28-SEP-1999; 990S-0156458.
PK 29-SEP-1999; 990S-0156596.
PK 04-OCT-1999; 990S-0157117.
PK 05-OCT-1999; 990S-0157753.
PK 06-OCT-1999; 990S-0157865.
PK 07-OCT-1999; 990S-0158029.
PK 08-OCT-1999; 990S-0158232.
PK 12-OCT-1999; 990S-0158369.
PK 13-OCT-1999; 990S-0159293.
PK 13-OCT-1999; 990S-0159294.
PK 13-OCT-1999; 990S-0159295.
PK 14-OCT-1999; 990S-0159329.
PK 14-OCT-1999; 990S-0159330.
PK 14-OCT-1999; 990S-0159331.
PK 14-OCT-1999; 990S-0159637.
PK 14-OCT-1999; 990S-0159638.
PK 18-OCT-1999; 990S-0159584.
PK 21-OCT-1999; 990S-0160741.
PK 21-OCT-1999; 990S-0160767.
PK 21-OCT-1999; 990S-0160768.
PK 21-OCT-1999; 990S-0160770.
PK 21-OCT-1999; 990S-0160814.
PK 21-OCT-1999; 990S-0160815.
PK 22-OCT-1999; 990S-0160980.
PK 22-OCT-1999; 990S-0160981.
PK 22-OCT-1999; 990S-0160989.
PK 25-OCT-1999; 990S-0161404.
PK 25-OCT-1999; 990S-0161406.
PK 25-OCT-1999; 990S-0161406.

```

```

PK 26-OCT-1999; 990S-0161359.
PK 26-OCT-1999; 990S-0161360.
PK 26-OCT-1999; 990S-0161361.
PK 26-OCT-1999; 990S-0161361.
PK 26-OCT-1999; 990S-0161361.
PK 26-OCT-1999; 990S-0161361.
PK 26-OCT-1999; 990S-0161361.
PK 26-OCT-1999; 990S-0161361.
PK 26-OCT-1999; 990S-0161361.
PK 26-OCT-1999; 990S-0161361.

```

Query Match 4.0%; Score 86; DB 21; Length 740;

Host Local 21; Conserved 34; Mismatch 115; Indels 120; Gaps 15;

```

QY 77 DAVGTHK-----KOVKKEKOC-----SSSHFASHTTCTGALHRTSSK 121
DB 16 dvve-dpyyqpaa pkyvynclvdcltracdploasgspvasgvyatldalvkerlanh 78
QY 122 YFNFTEDHKKVEALAKWASTHMSSTWTLKLTIVQIMQVYFDTCTVLAARAKVKA---- 177
DB 79 qvqclltauvavd-----lqysqkclhqlpsltpqpylqavvta 119
QY 178 -SPLA-KVW-----ICTRTTP-----PSMRKLLQDLVLRGDSHILRLQ 214
DB 120 lsl l l dclwyvclaat l pael l vradual l pael l dlsnvt l l kdlap l qdv l wldclq 178
QY 215 NLTDLHVNDRHNVARS--LKWGSLKLYSLHRTKLS- ---RNSLQCTMLSLDF 263
DB 179 s pldsl l kkvkvkvavrsqyllat l m-ayalcyudkclcl l dcl l vavvqql l dclq 236
QY 264 H-----LACGTSYGGGILV-----DE----- 280
DB 237 dsqsl l l l l pnpqkppvcl l wqcl l ncl l k l l aqg vewcl svtcl l cl l dcl l dcl 296
QY 281 --SNSVDOLKRLDFSANLTANQHSIMLCSEFPDSDSPNG 318
DB 297 l rslcl l vaaavrcvcl n svtcl dcl l vsgsqspdpvpsq 336

```

Search completed: April 9, 2002, 17:08:13

Job time: 592 sec

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138340.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140395.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145921.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151063.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154036.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155686.
PR 24-SEP-1999; 99US-0155659.
PR 26-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158569.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

ID AAW23594 standard; Protein: 3788 AA.
 XX AAW23594;
 AC 10-FEB-1998 (first entry)
 DT Murine Lystl long isoform.
 XX Lystl; mouse; lysosomal trafficking regulator;
 DE Chediak-Higashi syndrome; CH syndrome; autoimmune disease; tumour;
 KW vaccine; therapy; diagnosis.
 KW Mus musculus.
 OS WO9728262-A1.
 XX 07-AUG-1997.
 PN 31-JAN-1997; 97WO-US01748.
 PD 23-DEC-1996; 96US-0034346.
 XX 01-FEB-1996; 96US-0011146.
 PR 20-DEC-1996; 96US-0033599.
 XX (UYFL) UNIV FLORIDA.
 PA Barbosa-Alleyne MDES, Kingsmore SF;
 PI WPI; 1997-402616/37.
 DR N-PSDB; AAT74199.
 XX Mammalian lysosomal trafficking regulators LYST1, LYST1 and
 PT LYST2 - useful to diagnose Chediak-Higashi syndrome
 XX Claim 3; Page 101-103; 237pp; English.
 XX This protein comprises the longer isoform of murine lysosomal
 CC trafficking regulator Lyst1, a protein that regulates desgranulation
 CC of lysosomes, late endosomes and acidic secretory granules
 CC primarily in leukocytes. Its sequence was deduced from a Lyst1
 CC cDNA clone (AAT74199) that had been identified from a beige (bg)
 CC mutation critical region yeast artificial chromosome. Alternative
 CC splicing also yields a shorter isoform (see AAW23595) of Lyst1.
 CC Homologous human sequences (see AAW23596 and AAW23597) have been
 CC identified that are associated with Chediak-Higashi syndrome (CHS).
 CC Murine Lyst1 and human LYST1 polypeptides (see AAW21591-98) can be
 CC expressed in recombinant host cells for use in raising antibodies,
 CC in vaccines and in the development of therapeutic approaches to
 CC treatment of autoimmune diseases and certain types of tumours. The
 CC existence of an animal model of CHS with a similar genetic lesion
 CC will assist efforts to develop novel therapies for this disease.
 XX Sequence 1788 AA;
 SQ
 Query Match 5.0A; Score 88; DB 18; Length 3788;
 Best Local Similarity 18.8A; Pred. No. 43;
 Matches 81; Conservative 52; Mismatches 117; Indels 182; Gaps 17;
 QY 44 GSVVTVTDQTNVKTCLLFCNVEALPITYKNGDIDVRFHRLKIQVYKKETQGITSSGFAS 101
 Db 1849 cnglsmihqvlvkkc-----lqvfhllk----- 1872
 QY 104 LTFEGTIGAPIIPTSSKYNFETEDHKWVEALR-----VWAS----- 141
 Db 1873 tllegccgeevihvsehgefkldvshaiqlvklqellldwklnkacqgwwetlta 1932
 QY 142 -----TIMSPSWTLKLCQVPMQYEDLTCQLCK-----AE 173
 Db 1933 levllrvhqqqnlkqlnahvvhflitcqlvqhregqltmsprevcsvfklae 1992
 QY 174 VQGA-----SFLKAVDGTTR----- 189
 Db 174 VQGA-----SFLKAVDGTTR----- 189

Db 1993 vldspddlelltvlfnflavhptntvchmptnfvshldqkfkqkvqslaylrhs 2052
 QY 190 -----PPPSWRVLR-----QDLVLEGDLSHIHLQLNLTIDILVYDNHVVH 228
 Db 2053 ssggqafpspgflvispsaftaappetsssnlypqlmaaqmvrslp-----afptylp 2108
 QY 229 VARSCLKVGSFLRIYSLHTKLQSM---NSENQTMLSLEFHLHGCTSYGRGIRVLPESSNDV 285
 Db 2109 lraqkilaaslg-fsv-dkllqladanpekmlllrpya lk--tskoeafisscesaktv 2164
 QY 286 DOLKKDLESANLTAN-----QHSIDVICOSEPPDDSPNGVSL--RP---PCW 126
 Db 2165 cemea-llqahasagvscspfrprarvdhkvqleprsdtdspjdesyprtrpdmklgl 2223
 QY 327 SSVARSRLIAAS 338
 Db 2224 asfqsqstvas 2235
 RESULT 14
 AAG38741
 ID AAG38741 standard; Protein: 1002 AA.
 XX AAG38741;
 AC AAG38741;
 XX 18-OCT-2000 (first entry)
 DT Arabidopsis thaliana protein treatment. SEQ ID NO: 47839.
 DE Arabidopsis thaliana
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 PN EPI033405-A2.
 XX 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-0301439.
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 19-MAY-1999; 99US-0134768.
 PR 20-MAY-1999; 99US-0134941.
 PR 21-MAY-1999; 99US-0135124.
 PR 24-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.

PR	21-JUL-1999;	99US-0145088;
PR	22-JUL-1999;	99US-0145085;
PR	22-JUL-1999;	99US-0145087;
PR	22-JUL-1999;	99US-0145089;
PR	22-JUL-1999;	99US-0145192;
PR	23-JUL-1999;	99US-0145145;
PR	23-JUL-1999;	99US-0145218;
PR	23-JUL-1999;	99US-0145224;
PR	26-JUL-1999;	99US-0145276;
PR	27-JUL-1999;	99US-0145913;
PR	27-JUL-1999;	99US-0145918;
PR	27-JUL-1999;	99US-0145919;
PR	28-JUL-1999;	99US-0145951;
PR	02-AUG-1999;	99US-0146386;
PR	02-AUG-1999;	99US-0146388;
PR	02-AUG-1999;	99US-0146389;
PR	03-AUG-1999;	99US-0146738;
PR	04-AUG-1999;	99US-0147204;
PR	04-AUG-1999;	99US-0147300;
PR	05-AUG-1999;	99US-0147192;
PR	06-AUG-1999;	99US-0147260;
PR	06-AUG-1999;	99US-0147303;
PR	06-AUG-1999;	99US-0147416;
PR	09-AUG-1999;	99US-0147493;
PR	09-AUG-1999;	99US-0147935;
PR	10-AUG-1999;	99US-0148171;
PR	11-AUG-1999;	99US-0148319;
PR	12-AUG-1999;	99US-0148341;
PR	13-AUG-1999;	99US-0148565;
PR	13-AUG-1999;	99US-0148684;
PR	16-AUG-1999;	99US-0149368;
PR	17-AUG-1999;	99US-0149175;
PR	18-AUG-1999;	99US-0149426;
PR	20-AUG-1999;	99US-0149722;
PR	20-AUG-1999;	99US-0149723;
PR	20-AUG-1999;	99US-0149902;
PR	23-AUG-1999;	99US-0149930;
PR	25-AUG-1999;	99US-0150566;
PR	26-AUG-1999;	99US-0150884;
PR	26-AUG-1999;	99US-0150884;
PR	27-AUG-1999;	99US-0151065;
PR	27-AUG-1999;	99US-0151066;
PR	27-AUG-1999;	99US-0151080;
PR	30-AUG-1999;	99US-0151303;
PR	31-AUG-1999;	99US-0151438;
PR	01-SEP-1999;	99US-0151930;
PR	01-SEP-1999;	99US-0152363;
PR	07-SEP-1999;	99US-0153070;
PR	13-SEP-1999;	99US-0153758;
PR	15-SEP-1999;	99US-0154018;
PR	16-SEP-1999;	99US-0154039;
PR	20-SEP-1999;	99US-0154779;
PR	22-SEP-1999;	99US-0155139;
PR	23-SEP-1999;	99US-0155486;
PR	24-SEP-1999;	99US-0155659;
PR	28-SEP-1999;	99US-0156458;
PR	29-SEP-1999;	99US-0156596;
PR	04-OCT-1999;	99US-0157751;
PR	06-OCT-1999;	99US-0157765;
PR	07-OCT-1999;	99US-0158025;
PR	08-OCT-1999;	99US-0158232;
PR	12-OCT-1999;	99US-0158169;
PR	13-OCT-1999;	99US-0159291;
PR	13-OCT-1999;	99US-0159294;
PR	13-OCT-1999;	99US-0159295;
PR	14-OCT-1999;	99US-0159329;
PR	14-OCT-1999;	99US-0159330;
PR	14-OCT-1999;	99US-0159331;
PR	14-OCT-1999;	99US-0159637;
PR	14-OCT-1999;	99US-0159638;
PR	18-OCT-1999;	99US-0159584;
PR	21-OCT-1999;	99US-0160741;

[illegible]

RESULT 11

RESUL
AAG98362

Protein	Protein ID	Protein Accession
Protein 1	1	1
Protein 2	2	2
Protein 3	3	3
Protein 4	4	4
Protein 5	5	5
Protein 6	6	6
Protein 7	7	7
Protein 8	8	8
Protein 9	9	9
Protein 10	10	10
Protein 11	11	11
Protein 12	12	12
Protein 13	13	13
Protein 14	14	14
Protein 15	15	15
Protein 16	16	16
Protein 17	17	17
Protein 18	18	18
Protein 19	19	19
Protein 20	20	20
Protein 21	21	21
Protein 22	22	22
Protein 23	23	23
Protein 24	24	24
Protein 25	25	25
Protein 26	26	26
Protein 27	27	27
Protein 28	28	28
Protein 29	29	29
Protein 30	30	30
Protein 31	31	31
Protein 32	32	32
Protein 33	33	33
Protein 34	34	34
Protein 35	35	35
Protein 36	36	36
Protein 37	37	37
Protein 38	38	38
Protein 39	39	39
Protein 40	40	40
Protein 41	41	41
Protein 42	42	42
Protein 43	43	43
Protein 44	44	44
Protein 45	45	45
Protein 46	46	46
Protein 47	47	47
Protein 48	48	48
Protein 49	49	49
Protein 50	50	50
Protein 51	51	51
Protein 52	52	52
Protein 53	53	53
Protein 54	54	54
Protein 55	55	55
Protein 56	56	56
Protein 57	57	57
Protein 58	58	58
Protein 59	59	59
Protein 60	60	60
Protein 61	61	61
Protein 62	62	62
Protein 63	63	63
Protein 64	64	64
Protein 65	65	65
Protein 66	66	66
Protein 67	67	67
Protein 68	68	68
Protein 69	69	69
Protein 70	70	70
Protein 71	71	71
Protein 72	72	72
Protein 73	73	73
Protein 74	74	74
Protein 75	75	75
Protein 76	76	76
Protein 77	77	77
Protein 78	78	78
Protein 79	79	79
Protein 80	80	80
Protein 81	81	81
Protein 82	82	82
Protein 83	83	83
Protein 84	84	84
Protein 85	85	85
Protein 86	86	86
Protein 87	87	87
Protein 88	88	88
Protein 89	89	89
Protein 90	90	90
Protein 91	91	91
Protein 92	92	92
Protein 93	93	93
Protein 94	94	94
Protein 95	95	95
Protein 96	96	96
Protein 97	97	97
Protein 98	98	98
Protein 99	99	99
Protein 100	100	100

[illegible]

AC 98362;

XX
XXXXXX

DT 21-SEP-20

XX
XX

Dr. Escherich

XX
KW Escherichia coli; identification; proliferation; microorganism;
KW antimicrobial; antibacterial; antibiotic; aseptic therapy; diagnosis;
KW bacterial growth inhibition.

XX
XX
101 755 270 441

OS *Escherichia coli*.

[illegible]

PN WO200148209-A2.

XX

PD 05-JUL-2001

XX

PF 19-DEC-2000;

[illegible]

23-DEC-1999;

XX
BA
/EIT- \ EITDA DUADM INC

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140351.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 20-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159291.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159129.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159131.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.


```

Db 181 lkwdgttrtpfswrvliqdlvlegdlsghrlqnlctidilvydnhvnhvarsikvgsflr 240
QY 241 TYSLHTKLQSMNSENOTMLSLFHLHGTSYGRGIRVLPSNSDSDVQDLKKDLSEANLTAN 300
Db 241 Tyslhtklqsmnsenotmlslfhlhgtsygrgirvlpnsdsdvqdlkldlesanltan 300
QY 301 QHSDVICQSEPDSPNGVSL 321
Db 301 qhsdvicqsepdspngvsl 321

RESULT 3
AAB93478
ID AAB93478 standard; Protein: 634 AA.
XX AC AAB93478;
XX DT 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:12761.
XX Human; primer: detection; diagnosis; antisense therapy: gene therapy.
XX KW Human; primer: detection; diagnosis; antisense therapy: gene therapy.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PP 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isonai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Saitiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WP: 2001-318749/34.
XX primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX Claim 8: SEQ ID 12761; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dt primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB93893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification

```

```

CC of the present invention.
XX SQ Sequence 634 AA;
XX Query Match 93.5%; Score 1660; DB 22; Length 634;
XX Best Local Similarity 98.8%; Prod. No. 2.9a-171;
XX Matches 317; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 MSLVPATNYLYTPLNOLKGDTLVNVYGVKFFKPPVLSKGTDYCSVVTIVDOTNVALTCL 60
Db 1 mslvpatnylytplnolkgtlvnvvgvkvklfkppylskgtcsvvtlvdqlnvklcl 60
QY 61 LFSGNYEALPIIYKNGDIVRFHRLKTOVYKKEGTTSSGFSALTFEGTLCADIPRTSS 120
Db 61 lfsngyealpilykngdivrfhrlkiqvyykktggtssgfsaltfegtliqapilptss 120
QY 121 KYENFTTEDHKKVEALRVWASTHMSPSWTLKLCDDVOPMOYFDLTCOLLGKAEVDGASFL 180
Db 121 kyntfttedhkhmvealrvwasthmspswtllk lcdvqpmqyldlctqltqkaevdgsfl 180
QY 181 LKWDGTTRTPFWSRVLIQDLVLEGLDLSHRLHRIQNETIDILVYDNHVHVARSLKVGSLFR 240
Db 181 lkwdgttrtpfswrvliqdlvlegdlsghrlqnlctidilvydnhvnhvarsikvgsflr 240
QY 241 TYSLHTKLQSMNSENOTMLSLFHLHGTSYGRGIRVLPSNSDSDVQDLKKDLSEANLTAN 300
Db 241 Tyslhtklqsmnsenotmlslfhlhgtsygrgirvlpnsdsdvqdlkldlesanltan 300
QY 301 QHSDVICQSEPDSPNGVSL 321
Db 301 qhsdvicqsepdspngvsl 321

RESULT 4
AAB95120
ID AAB95120 standard; Protein: 634 AA.
XX AC AAB95120;
XX DT 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:17113.
XX Human; primer: detection; diagnosis; antisense therapy: gene therapy.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PP 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isonai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Saitiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WP: 2001-318749/34.
XX primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX Claim 8: SEQ ID 17113; 2537pp + CD ROM; English.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:08:10 ; Search time 58.75 seconds
(without alignments)
428.679 Million cell updates/sec

Title: US-09-816-248-15
Perfect score: 1776
Sequence: 1 MSLLPATNVIYTPNLQKLG.....LRPGWSVAKSRLLIAASTS 340

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*

1:	/SIDSB/qcdata/qcseq/qcseq/AA1980.DAT.*
2:	/SIDSB/qcdata/qcseq/qcseq/AA1981.DAT.*
3:	/SIDSB/qcdata/qcseq/qcseq/AA1982.DAT.*
4:	/SIDSB/qcdata/qcseq/qcseq/AA1983.DAT.*
5:	/SIDSB/qcdata/qcseq/qcseq/AA1984.DAT.*
6:	/SIDSB/qcdata/qcseq/qcseq/AA1985.DAT.*
7:	/SIDSB/qcdata/qcseq/qcseq/AA1986.DAT.*
8:	/SIDSB/qcdata/qcseq/qcseq/AA1987.DAT.*
9:	/SIDSB/qcdata/qcseq/qcseq/AA1988.DAT.*
10:	/SIDSB/qcdata/qcseq/qcseq/AA1989.DAT.*
11:	/SIDSB/qcdata/qcseq/qcseq/AA1990.DAT.*
12:	/SIDSB/qcdata/qcseq/qcseq/AA1991.DAT.*
13:	/SIDSB/qcdata/qcseq/qcseq/AA1992.DAT.*
14:	/SIDSB/qcdata/qcseq/qcseq/AA1993.DAT.*
15:	/SIDSB/qcdata/qcseq/qcseq/AA1994.DAT.*
16:	/SIDSB/qcdata/qcseq/qcseq/AA1995.DAT.*
17:	/SIDSB/qcdata/qcseq/qcseq/AA1996.DAT.*
18:	/SIDSB/qcdata/qcseq/qcseq/AA1997.DAT.*
19:	/SIDSB/qcdata/qcseq/qcseq/AA1998.DAT.*
20:	/SIDSB/qcdata/qcseq/qcseq/AA1999.DAT.*
21:	/SIDSB/qcdata/qcseq/qcseq/AA2000.DAT.*
22:	/SIDSB/qcdata/qcseq/qcseq/AA2001.DAT.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1660	93.5	634	22	AA040125
2	1660	93.5	634	22	AA092742
3	1660	93.5	634	22	AA093478
4	1660	93.5	634	22	AA095120
5	125	7.0	467	21	AA058505
6	103	5.8	291	21	AA167420
7	102.5	5.8	4393	22	AA031889
8	98	5.5	955	21	AA030852
9	98	5.5	1058	21	AA030851
10	98	5.5	1198	21	AA030850
11	93.5	5.3	400	22	AA098362

12	88.5	5.0	653	22	AA047250	Human PP7. Homo s
13	88	5.0	1788	18	AA023494	Murine Lysyl long
14	86.5	4.9	1002	21	AA038741	Arabidopsis thalia
15	86	4.8	740	21	AA042603	Arabidopsis thalia
16	86	4.8	790	21	AA042602	Arabidopsis thalia
17	86	4.8	899	21	AA042601	Arabidopsis thalia
18	85	4.8	510	20	AA022201	Human extracellular
19	85	4.8	510	21	AA054368	Protein encoded by
20	85	4.8	510	21	AA066646	Membrane-bound pro
21	85	4.8	510	22	AA03653	Human extracellular
22	85	4.8	510	22	AA065169	Human PRO698 (UNO3
23	85	4.8	510	22	AA050955	Human PRO698 prote
24	85	4.8	2285	20	AA098149	Bacillus subtilis
25	84	4.7	582	19	AA062996	Human truncated St
26	84	4.7	701	14	AA041335	84 kD ISGF-3alpha.
27	84	4.7	712	16	AA072079	Human STAT4. Hom
28	84	4.7	712	17	AA003170	Human STAT1-beta.
29	84	4.7	712	19	AA062995	Human STAT1-beta p
30	84	4.7	739	14	AA041334	91 kD ISGF-3alpha.
31	84	4.7	750	16	AA072078	Human STAT1. Hom
32	84	4.7	750	17	AA003168	Human STAT1-alpha.
33	84	4.7	750	19	AA062994	Human signal trans
34	84	4.7	750	22	AA019962	Human pancreatic c
35	84	4.7	786	21	AA054098	Porphorymonas qing
36	83.5	4.7	320	20	AA034350	Human b2 protein a
37	83.5	4.7	3801	18	AA031949	Arabidopsis thalia
38	83	4.7	467	21	AA023071	H. pylori GHP1161
39	83	4.7	508	19	AA098852	Arabidopsis thalia
40	83	4.7	536	21	AA031373	Arabidopsis thalia
41	83	4.7	537	21	AA023070	Arabidopsis thalia
42	83	4.7	557	21	AA023069	Arabidopsis thalia
43	83	4.7	606	21	AA031372	Arabidopsis thalia
44	83	4.7	623	21	AA025493	Eucalyptus grandis
45	83	4.7	626	21	AA031371	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AA040125

ID AA040125 standard; Protein: 634 AA.

XX

AA040125;

XX

DT 22-OCT-2001 (first entry)

XX

Human polypeptide SEQ ID NO 1270.

DE

XX

Human; non-tropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemo-lactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

KW Leukemia.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```

repeat_region      /rpl_family-"(TA)n"
                    35111..35131
repeat_region      /rpl_family-"AT-rich"
                    complement(35529..35823)
repeat_region      /rpl_family-"Alusg"
                    36436..36743
repeat_region      /rpl_family-"Alusp"
                    39015..39069
repeat_region      /rpl_family-"(TC)n"
                    complement(39134..39436)
repeat_region      /rpl_family-"MAKNA"
                    complement(39597..39733)
repeat_region      /rpl_family-"FLAM_C"
                    40838..40930
repeat_region      /rpl_family-"MIR"
                    complement(41206..41509)

Query Match,      7.5%: Score 97.2; DH 9; Length 16618;
Best Local Similarity 66.1%; Pred. No. 1.7e-15;
Matches 156; Conservative 0; Mismatches 78; Indels 2; Gaps 1;

QY 837 aagaagtaacatcagatcagatcgaactgaagaagagallagaatcgaatattgaagac 896
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145748 AGCAGTAAATATTTTTCATTGCACTATTTGCGATCTGATCTGCAATAAATGCAATAC 145807

QY 897 caatcagacatcagatgatactgtaatcagaacatcagacatcagacatcagacat 956
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145808 CCATAGAGCTTCCTTTTATTTTCGTTTATTTTATTTTATTTTATTTTTCAGATAGACT 145867

QY 957 ctcgctcgcctcgcagagctgaagtcgaagcagatcgcgcatactgaagcctccac 1016
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145868 CTCGCTCTGCCGCCAGCTGGAGTACAGTGGTCCGATCTCCGCTCAGTACACCCCTCCAC 145927

QY 1017 ctccatgaatcgaac--ttctcctcagctcagacctcccaagtaactgaatatarcaa 1070
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145928 CTCGCCGCTTCAGAGCAATCTTCTGCCCTTACGCCCTCCCAAGTACGTCGCAATACAGC 145983

```

Search completed: April 9, 2002, 23:35:53
 Job time: 1187 sec

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glende, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N., Hados, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lahoczky, J., Levine, R., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Melidrin, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Sougez, C., Spencer, B., Stange-thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trioglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.:
 Direct Submission
 Submitted (31-JUL-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 31, 2001 this sequence version replaced qt:14670134.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

 Center: Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIDR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

 Project Information
 Center project name: L18
 Center clone name: L18_E_18

We are submitting the entire project as phase 3. The T7 end overlaps accession number AC005235 (WICGR project L389) by 18.8 kilobases.

FEATURES	source	Location/Qualifiers
	1. .166518	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="17"	
	/map="17"	
	/clone_lib="RPC1-11 Human Male BAC"	
repeat_region	49. .345	
repeat_region	/rpt_family="AluY"	
repeat_region	1704. .2064	
repeat_region	/rpt_family="L2"	
repeat_region	complement(2079. .2385)	
repeat_region	/rpt_family="AluX"	
repeat_region	2365. .2756	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(3086. .3302)	
repeat_region	/rpt_family="AluS/x"	
repeat_region	complement(3375. .3499)	
repeat_region	/rpt_family="MIR"	
repeat_region	3662. .3704	
repeat_region	/rpt_family="(GGGA)n"	
repeat_region	4343. .4571	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(4685. .4991)	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(6025. .6194)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(6198. .6413)	
repeat_region	/rpt_family="MER20"	
repeat_region	complement(6505. .6811)	
repeat_region	/rpt_family="AluSx"	
repeat_region	7601. .7620	
repeat_region	/rpt_family="(CA)n"	
repeat_region	complement(8265. .8477)	
repeat_region	/rpt_family="MLT1J2"	
	complement(8490. .8683)	
	/rpt_family="MIR"	
repeat_region	8944. .8993	
repeat_region	/rpt_family="(TG)n"	
unsire	complement(9444. .9485)	
	/note="<30 qual single clone coverage"	
	complement(10463. .10532)	
repeat_region	/rpt_family="L3"	
repeat_region	complement(10696. .10829)	
repeat_region	/rpt_family="L2"	
repeat_region	complement(10942. .11071)	
repeat_region	/rpt_family="L2"	
repeat_region	complement(11559. .11717)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(11776. .12054)	
repeat_region	/rpt_family="AluSx"	
repeat_region	14340. .14492	
repeat_region	/rpt_family="MIR3"	
repeat_region	14999. .15060	
repeat_region	/rpt_family="MER41B"	
repeat_region	complement(15175. .15321)	
repeat_region	/rpt_family="MER5A"	
repeat_region	complement(15328. .15801)	
repeat_region	/rpt_family="LTR50"	
repeat_region	complement(15806. .15910)	
repeat_region	/rpt_family="MIR"	
repeat_region	16592. .16858	
repeat_region	/rpt_family="L2"	
repeat_region	17373. .17457	
repeat_region	/rpt_family="MER103"	
repeat_region	19279. .19576	
repeat_region	/rpt_family="L2"	
repeat_region	19778. .19830	
repeat_region	/rpt_family="L2"	
repeat_region	20330. .20626	
repeat_region	/rpt_family="AluSx"	
repeat_region	20857. .21030	
repeat_region	/rpt_family="MIR3"	
repeat_region	21878. .22012	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(24397. .24685)	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(25649. .25738)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(25854. .25951)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(26035. .26101)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(26382. .26554)	
repeat_region	/rpt_family="MIR"	
repeat_region	complement(26567. .26957)	
repeat_region	/rpt_family="L2"	
repeat_region	27040. .27345	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(28771. .29065)	
repeat_region	/rpt_family="AluSx"	
repeat_region	29405. .29592	
repeat_region	/rpt_family="MER5A"	
repeat_region	complement(29769. .29891)	
repeat_region	/rpt_family="MIR3"	
repeat_region	complement(30555. .30773)	
repeat_region	/rpt_family="L1PA5"	
repeat_region	30883. .31182	
repeat_region	/rpt_family="AluSx"	
repeat_region	complement(31357. .31916)	
repeat_region	/rpt_family="MLT1G1"	
repeat_region	complement(31946. .32059)	
repeat_region	/rpt_family="MIR"	
repeat_region	32480. .32765	
repeat_region	/rpt_family="AluJo"	
repeat_region	34083. .34337	
repeat_region	/rpt_family="MIR"	
repeat_region	34392. .34426	


```

misc_feature /chromosome="11"
misc_feature /clone="RP11-706A13"
misc_feature 1. 1325
misc_feature /note="assembly_name:Contig4"
misc_feature 1426. 2586
misc_feature /note="assembly_name:Contig5"
misc_feature 2687. 4338
misc_feature /note="assembly_name:Contig6"
misc_feature 4439. 5989
misc_feature /note="assembly_name:Contig7"
misc_feature 6090. 8269
misc_feature /note="assembly_name:Contig8"
misc_feature 8370. 10239
misc_feature /note="assembly_name:Contig9"
misc_feature 10340. 13484
misc_feature /note="assembly_name:Contig10"
misc_feature 13585. 18186
misc_feature /note="assembly_name:Contig11"
misc_feature 18287. 23764
misc_feature /note="assembly_name:Contig12"
misc_feature 23865. 27992
misc_feature /note="assembly_name:Contig13"
misc_feature 28093. 31755
misc_feature /note="assembly_name:Contig14"
misc_feature 31856. 36229
misc_feature /note="assembly_name:Contig15"
misc_feature 36330. 40157
misc_feature /note="assembly_name:Contig16"
misc_feature 40258. 45494
misc_feature /note="assembly_name:Contig17"
misc_feature 45595. 51920
misc_feature /note="assembly_name:Contig18"
misc_feature 52021. 58426
misc_feature /note="assembly_name:Contig19"
misc_feature 58527. 64891
misc_feature /note="assembly_name:Contig21"
misc_feature 64993. 72930
misc_feature /note="assembly_name:Contig22"
misc_feature 73031. 81807
misc_feature /note="assembly_name:Contig23"
misc_feature 81908. 89406
misc_feature /note="assembly_name:Contig24"
misc_feature /note="assembly_name:Contig25"
misc_feature 89507. 99182
misc_feature /note="assembly_name:Contig25"
misc_feature 99283. 111834
misc_feature /note="assembly_name:Contig26"
misc_feature 111935. 126258
misc_feature /note="assembly_name:Contig27"
misc_feature 126359. 134920
misc_feature /note="assembly_name:Contig28"
misc_feature 135021. 145335
misc_feature /note="assembly_name:Contig29"
misc_feature 145436. 162083
misc_feature /note="assembly_name:Contig30"
misc_feature 162184. 175617
misc_feature /note="assembly_name:Contig31"
misc_feature 175718. 190737
misc_feature /note="assembly_name:Contig32"

BANK COUNT 48045 a 46069 c 44860 g 49042 t 2723 others
ORIGIN

Query Match 7 6% Score 98.2; PH 2; Length 190737;
Best Local Similarity 85.8%; Pred. No. 9,4e-16;
Matches 109; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```

```

Db 175524 ACYCACCCCTACCTCCGCGACGTCAGACATCTCTGCTCAGCTCCCAAGTACCTGCA 175465
Cy 1063 attacag 1069
Db 175464 ACYACAGC 175458

RESULT 13
AC025588
LOCUS Homo sapiens chromosome 19, HAC C17-HSPC_16011 (HC288669),
DEFINITION complete sequence.
ACCESSION AC025588.1 GI:7229404
VERSION AC025588
KEYWORDS HPG.
SOURCE human.
ORGANISM Homo sapiens;
Chromosome: 19;
Genetic map: 145122 bp DNA
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 145122)
AUTHORS Lamerdin,J.E., McGready,P.M., Skowronski,E., Wisniallan,V.,
Burkhart-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Sullivan,W.,
Phan,H., Velasco,N., Do,L., Regalla,W., Terry,A., Brower,A.,
Garnes,J., Dargatzis,I., Erler,A., Christensen,M., Georgescu,A.,
Avila,J., Liu,S., Andreis,T., Frankel,M., Alix,C.,
Amico-Keller,G., Goettl,J., Duarte,S., Lucas,S., Bruce,R.,
Thomas,P., Quan,G., Kronmiller,B., Arias,A., Sanders,C., Wu,D.,
Nojan,M., Trogan,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
Sequence analysis of a 3-Mb region in 19q13.4
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 145122)
AUTHORS Lamerdin,J.E.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2000) JCB Joint Genome Institute, Lawrence
Livermore National Laboratory, Livermore, CA 94550
Map and sequence oriented from centromere to q-telomere. HAC
C17-HSPC_16011 is overlapped on the left by HC249609
(C17-HSPC_258823) and overlaps HC321036 (C17-HSPC_44484), AC005261)
on the right from bases 114293 to 145122 of this accession.
Additional chrom 19 map and sequence information are available at
http://www.bio.hut.fi/gen/dbp/genome/genome.html.
FEATURES
source 1. 145122
location:Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.4"
/clone="C17-HSPC_36011"
/clone_1lb="C17-HSPC"
/note="LLNL clone name: HC288669"
complement(385..562)
/rpl_family="L1MC5"
complement(775..931)
/rpl_family="L2"
complement(1187..1486)
/rpl_family="AlusX"
1690..1831
/rpl_family="L2"
1836..1960
/rpl_family="PIAM_C"
1961..2129
/rpl_family="(TA)n"
complement(2135..2503)
/rpl_family="L1M4"
2512..2805
/rpl_family="(TA)n"
2819..2938
/rpl_family="(TA)n"
3017..3126
/rpl_family="PIAM_A"
3127..3179
/rpl_family="AT-rich"

```

```

* 83651 83750: gap of unknown length
* 83751 93283: contig of 9533 bp in length
* 93284 93383: gap of unknown length
* 93384 102351: contig of 8968 bp in length
* 102352 102451: gap of unknown length
* 102452 129254: contig of 26803 bp in length
* 129255 129354: gap of unknown length
* 129355 176697: contig of 47343 bp in length.
FEATURES
    Location/Qualifiers
        1..176697
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="11"
        /clone="RP11-655D7"
BASE COUNT 47001 a 41857 c 40829 g 44654 t 2356 others
ORIGIN

Query Match 7.6%; Score 98.2; DB 2; Length 176697;
Best Local Similarity 85.8%; Pred. No. 9.4e-16;
Matches 109; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 943 ttccaaatggaatcgttcgttcctccagcagctgagattcaatgagcagctcgcgc 1002
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175089 TTTTGAGATGGAGCTCGCTCTCTCGCCGAGCTGCGATGCGATGATCTTGGCTC 175030

QY 1003 attgagctccacctcgttcgattcaagcttcctcctcagcctcccaagtagctgg 1062
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175029 ACTGCCAGCCTTACCTCCGAGCTCAAGATCTCTCGCTCGAGCTCCAGTAGCTGGA 174970

QY 1063 attacaq 1069
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174969 ACTACAG 174963

RESULT 12
AC019143/C
LOCUS AC019143 190737 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-706A11, WORKING DRAFT
SEQUENCE 28 unordered pieces.
AC019143
AC019143.3 GI:7715065
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 190737)
    Waterston,R.H.
    The sequence of Homo sapiens clone
    Unpublished
    2 (bases 1 to 190737)
    Waterston,R.H.
    Direct Submission
    Submitted (30-DEC-1999) Genome Sequencing Center, Washington
    University School of Medicine, 4444 Forest Park Parkway, St. Louis,
    MO 63108, USA
    On May 7, 2000 this sequence version replaced gi:7024180.
COMMENT
    ----- Genome Center -----
    Center: Washington University Genome Sequencing Center
    Center code: WUGSC
    Web site: http://genome.wustl.edu/gsc/index.shtml
    ----- Project Information -----
    Center project name: H_RH0706A11
    ----- Summary Statistics -----
    Sequencing vector: M13; 78%
    Chemistry: Dye-terminator Big Dye; 22% of reads
    Assembly program: Phrap; version 0.990319
    Consensus quality: 167325 bases at least Q40
    Consensus quality: 175317 bases at least Q30

```

Consensus quality: 180039 bases at least Q20
 Insert size: 193000; agarose-gel
 Insert size: 188037; sum-of-contigs
 Quality coverage: 3.04 in Q20 bases; agarose-gel
 Quality coverage: 3.18 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
 consists of 28 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1325: contig of 1325 bp in length
 1425: gap of unknown length
 1426 2586: contig of 1161 bp in length
 2587 2687: gap of unknown length
 2688 4438: contig of 1652 bp in length
 4439 5989: contig of 1551 bp in length
 5990 8269: gap of unknown length
 8270 8369: contig of 2180 bp in length
 8370 10239: contig of 1870 bp in length
 10240 10339: gap of unknown length
 10340 13484: contig of 3145 bp in length
 13485 13585: gap of unknown length
 13586 18187: contig of 4602 bp in length
 18188 18286: gap of unknown length
 18287 23764: contig of 5478 bp in length
 23765 23864: gap of unknown length
 23865 27992: contig of 4128 bp in length
 27993 28092: gap of unknown length
 28093 31755: contig of 3763 bp in length
 31756 31855: gap of unknown length
 31856 36229: contig of 4374 bp in length
 36230 36329: gap of unknown length
 36330 40157: contig of 3828 bp in length
 40158 40257: gap of unknown length
 40258 45494: contig of 5237 bp in length
 45495 45594: gap of unknown length
 45595 51920: contig of 6326 bp in length
 51921 52020: gap of unknown length
 52021 58426: contig of 6406 bp in length
 58427 58526: gap of unknown length
 58527 64892: contig of 6366 bp in length
 64893 64992: gap of unknown length
 64993 72930: contig of 7938 bp in length
 72931 81807: contig of 8777 bp in length
 81808 89406: contig of 7499 bp in length
 89407 89506: gap of unknown length
 89507 93182: contig of 9676 bp in length
 93183 93282: gap of unknown length
 93283 111814: contig of 12552 bp in length
 111815 111934: gap of unknown length
 111935 126258: contig of 14324 bp in length
 126259 126358: gap of unknown length
 126359 134920: contig of 8562 bp in length
 134921 135020: gap of unknown length
 135021 145335: contig of 10315 bp in length
 145336 145435: gap of unknown length
 145436 162083: contig of 16648 bp in length
 162084 162183: gap of unknown length
 162184 175617: contig of 13434 bp in length
 175618 175717: gap of unknown length
 175718 190737: contig of 15020 bp in length.

Location/Qualifiers
 1..190737
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

FEATURES
 source

COMMENT

Direct Submission
Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis
MO 63108, USA
On Mar 13, 2000 this sequence version replaced gi:7024164.

* NOTE: This is a working draft sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* As soon as it is available and the accession number will
* be preserved.

misc_feature
/c/one="RP1-255N19"
1. 196900

BASE COUNT	54333	a	41365	c	43067	g	58135	t
ORIGIN								

Query Match	7.68;	Score 98.6;	bits 2;	Length 196900;
Best Local Similarity	76.68;	Pred. No. 7.3e-16;		
Matches 134; Conservative	0;	Mismatches 39;	Indels 2;	Gaps 1

901 cagcaltcagatqitaltcltclcaatcaguarct'gacacacagct'it'ccuat'qqacl'ct'cg 966

Db 7588 CAGAAATTATATTTTATCTTCAGATACCATATATATCTTTTTTTTTCAGATGCGACTTTC 7647

[illegible]

1070

[illegible]

RESULT 11
AC021170/c

LOCUS	AC021170	1/869/ bp	NO	07-09-2008
DEFINITION	Homo sapiens chromosome 11 clone RP11-655D7, WORKING DRAFT			

ACCESSION NO. AC021170

KEYWORDS: HPG; HTGS_PHASE1; HPGS_DRA1², human source

ORGANISM Homo sapiens
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi

REFERENCE 1 (bases 1 to 176697)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

Authors	Waterston, R.H.
Title	The sequence of Homo sapiens clon-

REFERENCE	2 (bases 1 to 176697)
AUTHORS	Waterston, D. H.

```

Direct Submission
Submitted (14-Jan-2000) Genome Sequencing Center, Washin
MO University School of Medicine, 4444 Forest Park Parkway,
MO 63108, USA
On Mar 13, 2000 this sequence version replaced gi:702416

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H_MNH065D07
Summary Statistics
Sequencing vector: Plasmid; 100%
Chemistry: Dye-primed; 100% of reads
Assembly program: Phrap; version 0.99019
Consensus quality: 164861 bases at least Q40
Consensus quality: 164822 bases at least Q40
Consensus quality: 170901 bases at least Q20
Insert size: 164000; 1p
Insert size: 174193; sum-of-contigs
Quality coverage: 3.57 in Q20 bases; agatasc-1p
Quality coverage: 4.02 in Q20 bases; sum-of-contigs

```

NOTE: This is a "working draft" sequence. It currently consists of 24 contigs. The true order in this sequence record is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of Ns, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	1086: contig of 1086 bp in length
1087	gap of unknown length
1187	1186: contig of 3485 bp in length
4672	4671: gap of unknown length
4772	4771: gap of unknown length
7903	7902: contig of 5131 bp in length
7903	8002: gap of unknown length
8003	11020: contig of 3018 bp in length
11021	11120: gap of unknown length
11121	14825: contig of 3705 bp in length
14826	14825: gap of unknown length
14926	17464: gap of unknown length
17465	17564: contig of 2539 bp in length
17565	17564: gap of unknown length
21884	21883: contig of 4319 bp in length
21884	21983: gap of unknown length
25335	25335: contig of 3352 bp in length
25336	25435: gap of unknown length
25436	28938: contig of 3503 bp in length
28939	29038: gap of unknown length
29039	32405: contig of 3367 bp in length
32406	32505: gap of unknown length
32506	36284: contig of 3779 bp in length
36285	36384: gap of unknown length
36385	39620: contig of 3136 bp in length
39621	39620: gap of unknown length
39621	43360: contig of 3940 bp in length
43361	43360: gap of unknown length
43361	44604: gap of unknown length
48581	48581: contig of 4921 bp in length
48582	48681: gap of unknown length
48682	52884: contig of 4203 bp in length
52885	52984: gap of unknown length
52985	52984: contig of 4450 bp in length
52985	57334: gap of unknown length
57335	57334: gap of unknown length
57335	62596: contig of 5062 bp in length
62597	62596: gap of unknown length
62697	69908: contig of 6712 bp in length
69409	69608: gap of unknown length
69509	76036: contig of 6528 bp in length
76037	76036: gap of unknown length
76137	83550: contig of 7514 bp in length

- * NOTE: This is a 'working draft' sequence. It currently
- * consists of 3 contigs. The true order of the pieces
- * is not known and their order in this sequence record is

COMMENT

AC074366.5 GI:15144343
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVE[IN].
human.
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 196900)
Waterston, R.H.
The sequence of Homo sapiens clone:
Unpublished
2 (bases 1 to 196900)
Waterston, R.H.
Direct Submission
Submitted (29-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 9, 2001 this sequence version replaced gi:14626386.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/asg/index.shtml>
----- Project Information -----
Center project name: H NH0255N19

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

FEATURES

SOURCE

Location/Qualifiers

```

1..74155
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSHB-358P8"
/chromosome="X"
/clone_lib="Genome Systems Human BAC Library"
/map="Xp22"

1..206
/note="overlap bases 137066 to 137270 in AC004552."
complement(282..561)
/rpl_family="Alusx"
878..1054
/rpl_family="MIR"
complement(1783..2080)
/rpl_family="AluJb"
complement(2087..2380)
/rpl_family="Alusx"
complement(2384..2425)
/rpl_family="L2"
complement(2924..3108)
/rpl_family="L1MC4"
3138..3439
/rpl_family="Alusx"
3808..4265
/note="Region: Similar to M48206, YV22P03, r1 Sources: fetal liver spleen INPLS Homo sapiens cDNA clone 243437 5'"
4150..4291
/rpl_family="MLT1G"
complement(4703..4986)
/rpl_family="AluJo"
5404..5576
/rpl_family="L1PA16"
complement(5425..5939)
/note="Region: N39012 YV22P03.s1 Sources: fetal liver spleen INPLS Homo sapiens cDNA clone 243437 3'"
5987..6096
/rpl_family="AluSp/4"
6479..6785
/rpl_family="Alusx"
7475..7776
/rpl_family="Alusx"
complement(8221..8347)
/rpl_family="FLAM_A"
complement(8989..9062)
/rpl_family="L1MC4"
9113..9370
/rpl_family="Alusx"
complement(9374..9745)
/rpl_family="MER31A"
9746..10021
/rpl_family="AluJb"
complement(10022..10069)
/rpl_family="MER31A"
complement(10249..10269)
/rpl_family="AT-rich"
10297..10378
/rpl_family="AluJb"
10379..10669
/rpl_family="MER47H"
10792..11085
/rpl_family="Alusx"
/rpl_family="AluY"
complement(11793..11925)
/rpl_family="MIR"
join(12813..12955,16901..16972,21399..21583,23281..23370,37005..37082,44270..44346,45646..45747,46148..46258,49729..49842,52001..52497)
/genome="Homo sapiens pyruvate dehydrogenase kinase isoform 3"

```

```

repeat_region
/rpl_family="MER31B"
complement(14206..14403)
/rpl_family="L1MC4"
complement(14366..14460)
/rpl_family="(CA)n"
complement(14469..14724)
/rpl_family="L1MC4"
complement(14852..15151)
/rpl_family="AluJ"
complement(15355..15647)
/rpl_family="Alusx"
complement(17724..17814)
/rpl_family="MIR"
complement(18183..18506)
/rpl_family="MER2"
19504..19803
/rpl_family="Alusx"
20670..20933
/rpl_family="Alusq"
complement(24745..24860)
/rpl_family="L1MH6"
complement(24881..25180)
/rpl_family="Alusq"
complement(25183..25492)
/rpl_family="L1MH7"
25494..25640
/rpl_family="MIR"
complement(25752..25775)
/rpl_family="AT-rich"
complement(26089..26467)
/rpl_family="MSTA"
complement(26788..27088)
/rpl_family="Alusx"
27562..27637
/rpl_family="(CA)n"
29638..29936
/rpl_family="Alusq"
30636..30848
/rpl_family="Alusq/x"
complement(31901..31935)
/rpl_family="AT-rich"
32878..32956
/rpl_family="MER5A"
34392..34865
/rpl_family="L1MH7"
complement(34871..35166)
/rpl_family="Alusq"
35171..35227
/rpl_family="L1MH8"
complement(35295..35593)
/rpl_family="Alusx"
36273..36572
/rpl_family="Alusx"
46594..46775
/rpl_family="AluJo"
36741..36982
/standard_name="dbSTS:47229, AFM189x5, Chr. X, Homo sapiens"
36891..36954
/rpl_family="(CA)n"
37296..37428
/rpl_family="L1P5"
complement(37440..37741)
/rpl_family="Alusx"
37983..38280
/rpl_family="Alusx"
38284..38428
/rpl_family="L1MH"
39065..39370

```

	misc_feature	vp94d10.r1*	14398..14533	
	Query Match	12.7%, Score 164.4; DB 9;	Length 154950;	
	Best Local Similarity	96.6%; Ident. No. 2,50-11;		
	Matches 168; Conservative	0; Mismatches 6; Indels	0; Gaps	0;
OY	1122	tacatcttttcaatacattccatgatttgtagagagagagaccacattgtgcatttcbaaaca	1181	
	Db	6844 TACACTACTCTTCAAGATTCATGCTATTGTGCAGAGCATATATATGCCATTTTGAACA	6903	
OY	1182	aagaagctctcaaacacabcccgaacctggagaaacttgaqcttcaaaqccccaaaact	1241	
	Db	6904 AAAAGCCTCGCCACAAATACGGATCCGACCAAAATTTGAGGTCAATPMAAGCCACAGAAGCT	6963	
OY	1242	atttcagtcttttaaatctcatgccccaatatgtcatacttctcgcaagaagttcc	1295	
	Db	6964 ATTTCAGTCTCTTAAACTTCATGCGCCVAAAATGTCATGTCGTGAGATPTTG	7017	
RESULT	8			
AC0004656	LOCUS			
DEFINITION	Homo sapiens Xp22 HAC USH-35BP8 (Genome Systems Human BAC library)	PRT	02-JUL-1998	
ACCESSION	complete sequence.			
VERSION	AC0004656			
KEYWORDS	AC0004656.1 GI:3253114			
SOURCE	HTG.			
ORGANISM	human.			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 74155) Muzny,D., Aenson,A.D., Adams,C., Brandage,E., Bunae,G., Carvelli,K., Chacko,J., Chen,Y., Di,W., Ding,Y., Duan,S., Durbin,J., Foreman,J., Ganesh,R., Garcia,C., Goodman,M., Correll,J.H., Haywood,M., Hernandez,J., Jackson,L., Jin,S., Kampal,I.R., Karpathy,S., Kovar,C., Leal,B., Li,Y., Liechafte,O., Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L., Rashid,N.D., Rowland,K., Savage,L., Scherpf,S.E., Shen,H., Simon,M., Stoval,K., Tilms,K.M., Todd,J., Vo,O., Williamson,A., Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.			
TITLE	Direct Submission			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 74155)			
AUTHORS	Mortley,K.C.			
TITLE	Direct Submission			
JOURNAL	Submitted (02-MAY-1998) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 74155)			
REFERENCE	Mortley,K.C.			
AUTHORS	Direct Submission			
TITLE	Submitted (02-JUL-1998) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
JOURNAL	Baylor Plaza, Houston, TX 77030, USA			
COMMENT	On Jun 25, 1998 This sequence version replaced gi:3242666. Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.			
	The repeat regions shown were identified using RepeatMasker by Adrian Smith.			
	Sequence similarities were identified using Powerblast by Jinghui Zhang.			

```

Db 334 -----GCTCTCGATCAGTATCATATTACGAG 359
Gy 1096 gtaaaagatgacacagctatctctcaataactacagatcactatatttgaagag 1155
|||||
Db 360 GPAGAAGATCTCAAGAGCTATCTCTACAAATCTTACAGATCATCATTTTCAGAGG 419
Gy 1156 acaccactatgacgcttttgaagacaaagatcctcaacataccgacacacaa 1215
|||||
Db 420 ACACCACTATCTCCATTTTGAACAAAAGCTTCTCAACAATACCGATCCGACGACAAA 479
Gy 1216 ttaagctatagaccacagacatttcaatctctttaaacttcttcttcaatag 1275
|||||
Db 480 TTGAGCTCATATAGCCCAAGACAGATATTTCAGTCTGTAAATCTCATTTGCTTAATCT 539
Gy 1276 catttactgcaaaatttccaca 1298
|||||
Db 540 CATTTCTGCAAGAGCTTCACA 562

RESULT 7
AC004925 154959 bp DNA PRI 21-DEC-1999
DEFINITION Homo sapiens PAC clone RP5-907C10 from 7q31-q32, complete
sequence.
AC004925 AC004925
VERSION AC004925.1 GI:4156174
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 154959)
AUTHORS Jones, K., Tin-Wollam, A. and Koppeler, D.
TITLE The sequence of Homo sapiens PAC clone RP5-907C10
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 154959)
AUTHORS Waterston, K.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
3 (bases 1 to 154959)
AUTHORS Waterston, K.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
4 (bases 1 to 154959)
AUTHORS Waterston, K.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
5 (bases 1 to 154959)
AUTHORS Waterston, K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 14, 1999 this sequence version replaced gi:3213072.
----- Genome Center
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/qsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_00907C10

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:

all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the ENIGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.geni.nih.gov/DIC/GRN/CH7>, send <mailto:genome@geni.nih.gov>, or see <http://genome.wustl.edu/qsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RP5-5, prepared by Peter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Tomblin et al., Nature Genetics 6:84-9 (1994). The library is from one male donor. The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Peter de Jong.

VECTOR:

PCPAC2

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP5-907C10; actual end is at 154959 of RP5-907C10.

The sequence from 4689 to 5920 is from PCR product from PAC RP5-907C10 DNA.

FEATURES:

SOURCE

1..154959
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q31-q32"
/clone="RP5-907C10"
/clone_lib="RP5-5"
66..259

repeat_region

731..752
/rpt_family="L1"

repeat_region

1128..1429
/rpt_family="Alu"

repeat_region

1652..1791
/rpt_family="L1"

repeat_region

1810..1936
/rpt_family="L1"

repeat_region

1949..2296
/rpt_family="L1"

repeat_region

2297..2595
/rpt_family="Alu"

repeat_region

2596..2760
/rpt_family="L1"

gene

<2816..25850
/gene="WUGSC:H_030907C10.1"

CDS

join(<2816..2870,6849..7005,8634..8839,14398..14533,20470..20558,22507..22598,24455..24560,25738..25850)
/gene="WUGSC:H_030907C10.1"
/note="supported by human ESTs H23395 (NID:q892090) and AAL26363 (NID:q1687976) mouse ESTs W83982 (NID:q1394952) and AAT17633 (NID:q2729907), and GenScan; H_030907C10.1"
/codon_start=1
/evidence=not_experimental
/protein_id="AA08852.1"
/db_xref="GI:4176476"

/translation="SCSVSLYVERCCOLSATILTDHOYLERTPLCALLKKAPOOYK
HAKURSKPRRLFOSVKALHCPKCHILLOEVPHEGDLIIIFODGATKTPDKIQTSLY
DSK1WTTNKGKRVAVHFVHFNILPSNECLLLIEGTLSEIKLSKFNPSVPLV
SCHIEDLELDSAPFLIOCTIHHYGCCKCSLSRLSNLSVDKTSIPLSSVAFALG
VPLVYFVMTTLDKTCGLVLAFLMDSKFFQIPASEVLMDDDLQKSDVMIMMPCPI

Db 1388 -----GCTCTGGATCA 1398

QY 1081 qtaacataacaggaagaaatqtaacacgaatctactacataactaacacacac 1140
 |||||

Db 1399 GTATCATTTATACGAGCTAGAAAGATGTCAGACGATATCTGCTACAAATCTTACAAATCAT 1458
 |||||

QY 1141 gattatcttggaagagacacacataatgacattcttgaaagaaagctctcaaaacac 1200
 |||||

Db 1459 CAGTATTTGAGAGGACACACATATGTCGCAATTTGAAACAAAGCTCTCCACACAAVAC 1518
 |||||

QY 1201 cgaaccagagaaatgaatgaatcacaataagcagaagacatactcaatctgttaacac 1260
 |||||

Db 1519 CGCATCCGAGCAAAATGAGCGCATATTAAGCCAGAAAGCATTTTCACGCTGTAAACTT 1578
 |||||

QY 1261 gattgcctaaatqtaactatgctatgctgaagaatctcaca 1398
 |||||

Db 1579 CATTGCCCTAATGTCATTTGCTCCAGACAACTTCCACA 1616
 |||||

RESULT 4

LOCUS AK001230 2631 bp mRNA PRI 22-FEB-2000

DEFINITION Homo sapiens CDNA FLJ10368 fls, clone NT2RM2001544.

ACCESSION AK001230

VERSION AK001230.1 GI:7022354

KEYWORDS oligo capping; fls (full insert, sequence);

SOURCE Homo sapiens keratocarcinoma cell_line:NT2 CDNA to mRNA,
 clone_lib:NT2RM2 clone:NT2RM2001544.

ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Hominidae: Homo.
 1 (sites)

AUTHORS Isoqali,T., Ota,T., Hayashi,K., Sugano,S., Shitahara,A., Sudo,H.,
 Nishikawa,T., Nagai,K., Sugano,S., Shitahara,A., Sudo,H.,
 Magatsuma,M., Hosofuji,T., Kaku,Y., Kodaira,H., Kondo,H.,
 Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
 Ono,Y., Takiguchi,S., Matanabe,S., Kimura,K., Mueckli,K.,
 Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
 Nakamura,Y., Nagahori,K., Masino,Y., Ninomiya,K. and Iwayanagi,T.

TITLE NED0 human CDNA sequencing project.

JOURNAL Unpublished (2000)

REFERENCE 2 (bases 1 to 2631)

AUTHORS Isoqali,T. and Otsuki,T.

TITLE Direct Submission

JOURNAL Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases, Takao
 Isoqali, Helix Research Institute, Genomics Laboratory, 1512-1 Yama,
 Kisarazu, Chiba 2592-0812, Japan (E-mail:genomics@hrl.co.jp,
 Tel:81-48-52-1051, Fax:81-48-52-3952)

COMMENT NED0 human CDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan; CDNA full insert
 sequencing; Research Association for Biotechnology; CDNA library
 construction, 5'- & 3'- and one pass sequencing and clone selection;
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.

FEATURES

source

1..2631

location/Qualifiers

organism="Homo sapiens"

db_xref="taxon:9606"

cell_line="NT2"

cell_type="keratocarcinoma"

clone="NT2RM2001544"

clone_lib="NT2RM2"

note="cloning vector: pME18SFL3; mRNA from uninduced NT2
 neuronal precursor cells."

24..1928

note="unlabeled protein product"

protein_id="BA091568.1"

cds_start=1

cds_end="61:7022355"

translation="MSLVATATVTPPLNDLKCTTVNVGVKPKRPYLSKCTGYG
 SVVTVDFGVKLTCLFSCNVALPIYKNGDILVPHRLTGVYKKEFGVLTSSGPA
 SLTECTGLADILPRTSSKYFNFTEDHKVVALRWASVTHMSWTLKLCIDVQPMO

VFDTTCQILAKAVVPCASPLIAKVMGCTRTTPSPRVLLIDLVLEEDLSHILRLQNLTL
 DILVYDNIHVHARSILKVGSPFLRYSLSHTXIKOSMNSENQPMLSLEHLLGCTSYGGR
 VLPSNSDNDVILAKKIDLESANLITAMHOSVDPICQSPDPSPPSSGSAVLEVERCOOLA
 TILDLQLEERFLALILKOKAPQOYRISAKLRSYKPRRLPOSVLRHICRKLLOEVP
 HEEDLDLITPDCACTKTPVYKLTQNTSYRDXKIMTKNOKIKRKAIVHVRNKGILTPSNE
 CLILITRGTTLSERCKLSNKNFNSYIPVRSHEEDLILDSAPPLIGCTTHHYCKKCSS
 LKLSIONLSLVDRFTSMIPSSVAVALATVPLGVVFWVFTLIDKTVLEAVIMDSRFP
 QIIPASVLEMDDDIQRKSVDMIMDEKPHGIRIDAVVPLKECFIKSVNVTGCTNOICVQI
 FDTTVAEDVIT"

BASE COUNT 826 A 471 C 491 T 841 G

ORIGIN

Query Match 80.74; Score 1048; DB 9; Length 2631;

Best local similarity 90.4%; proc. No. 7, 6e-268;

Matches 1178; Conservative 0; Mismatches 0; Indels 120; Gaps 1;

QY 1 atgctcttggtccagacgaatctatctctatagacccctatcatcaatctatgct 60
 |||||

Db 24 ATGCTCTTGTGCTGCAGACAAATTAATATACCCCTCGAATCAACTTAAGAGTGCT 83
 |||||

QY 61 acaatctgcaatctctatctatgctatgctatctctctatagcccaatactcaagaaaga 120
 |||||

Db 84 ACAATTCGCAATGCTATGCTGTTGCAAGCTCTTTAAGCCCCCATATGTAACCAAGCA 143
 |||||

QY 121 actgaatctgctcgaatctgcaatctctgcaatctctgcaatctctgcaatctgctg 180
 |||||

Db 144 ACTGATTAATCTCAGTTGTAATGTTGTCAGACAGCAAAATGTAACATTAACCTGCTG 203
 |||||

QY 181 cctctctgagaaactctgaaagcctctgaaatctctctctctctctctctctctctct 240
 |||||

Db 204 CTTCTTACGCAAGCAATGAGCCCTTCGANTATTTATTAAGATGAGATATGTTCCG 263
 |||||

QY 241 ctctcagagctgaagatctcaatctataaaagagagctcaaggtatctcaacatctgagc 300
 |||||

Db 264 TTTTCACAGCTGAGAGATTCAGATATTAATAAAGAGAGCTGAGCTATGACAGCTGCGC 323
 |||||

QY 301 ctctcagatctctgctctgagagagctctgagagcctctctctctctctctctctct 360
 |||||

Db 324 TTTGCAATCTTGAAGCTTGAAGAGATTTGAGAGCTCTTGAATGAGCTGCACTTCAAGC 383
 |||||

QY 361 aagatct 420
 |||||

Db 384 AAGTATTTTAACTGATCTATGAGAGACCAAAATGTAACAGCTTACGCTTGGCGCA 443
 |||||

QY 421 tctatctcatct 480
 |||||

Db 444 TCTATCTATATGCTGACGCTCTTGAATGATTAATAATTTGATGATGATGATGATG 503
 |||||

QY 481 cactctgagct 540
 |||||

Db 504 TATTTTGAAGCTGATTTGATGATCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 563
 |||||

QY 541 ctcaagagct 600
 |||||

Db 564 CTAAAGCTATGAGATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 623
 |||||

QY 601 ct 660
 |||||

Db 624 CTCTCTCTCTGAGCTGATTTAAGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 683
 |||||

QY 661 ctgagctcagagct 720
 |||||

Db 684 TTACGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 743
 |||||

QY 721 atct 780
 |||||

Db 744 ATGATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 803
 |||||

QY 781 ctgaagct 840
 |||||

Db 804 TTAAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 863
 |||||


```

QY 112 APII-----PRISSEYFNFTIEDHKMVEAL-----RWASTHM 141
D 133 IPVYPASAOVALLPEKOCESKLLKEHHQVSYLYNVIDKYDVPDEAEYOORVKVSLNV 192
QY 145 SPSNTLLKCDVORVDFLTCL-----LGKAEV-----DGASFLLKVDGTGTRIP 190
D 193 KDFSLK--DVTGNYFDLQVAKDPYDEMGRIIYLSYDENTDLFFHYHTEGVRDL 250
QY 191 FPSWR---VLIODLVLEGLSH--IHLQNLITIDLYVDNHHVAV--SLKVGSLFLIYS 243
D 251 ASAAPADAYEDNPNPEAAQHPWGPYGRKRIIOISYDAHADFIKAGVSAGMWSLRN 310
QY 244 LHTK 247
D 311 VQVK 314

RESULT 10
QY3702 ID Q93702 PRELIMINARY: PRI: 587 AA.
AC Q93702
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DI 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE F39H2.1 PROTEIN.
GN F39H2.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
PP SEQUENCE FROM N.A.
RA Kershaw J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
PP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Haynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen P.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston P.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.
RL Nature 358:32-38(1994).
DR EMBL: 281080; CAB03086.1;
SQ SEQUENCE 587 AA: 55602 MW: CF8A55F14188446F CRC64;

```

```

Query Match 5.18: Score 109; DB 5: Length 587;
Best Local Similarity 22.98; Pred. No. 0.31; Mismatches 60; Indels 89; Gaps 10;
Matches 51; Conservative 24;

QY 122 YNFTIEDHKMVE-----ALRWASTHMSPSNTLLKLCVQPMQYFDLT-----COL-- 158
D 71 YFDVLAQVHSYVETKNGSWTLRWRAQKFGPEASKEK----RKNLPHVIENTFKCYIVP 126
QY 169 -----LGKA-EVDGASFLKVDGTGTRIPPSNRVLIODLVLEGLSHIHLQNLITIDLY 22?
D 127 PDPRIGRAVEESGREFL-----IETDLY 149
QY 224 DNHHVAVSLKVGSLFLIYSLH-----TKLOSMNSNOTMLSEFLHFGTISYGRGIRV 277
D 150 DHRPIGIEKLRSGDFVALONVHAASVGLTEMQVHLG-----GGQATNRGISK 196
QY 278 LPES--NSVDOLKLDLESA-----NLTANQSHDVICOSE 310
D 197 VPVDFRNEAFOIFKRKRVESVLEAVTDCDNFIEFOCKENNVQSE 239

```

```

RESULT 11
QY4E4 ID QY4E4 PRELIMINARY: PRI: 1044 AA.
AC QY4E4
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DI 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KIAA0580 PROTEIN (FRAGMENT).
GN KIAA0580.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
PP SEQUENCE FROM N.A.
PC TIISUE=BRAIN;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RI "Prediction of the coding sequences of unidentified human genes. IX.
RI The complete sequences of 100 new cDNA clones from brain which can
RI code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
DR ENBL: AB011152; BAA25506.1;
DR InterPro: IPR000198; RhoGAP.
DR InterPro: IPR001164; Znf_GCS.
DR InterPro: IPR001849; PH.
DR Pfam: PF00169; PH; 4.
DR Pfam: PF00620; RhoGAP; 1.
DR Pfam: PF01412; ArfGAP; 1.
DR PRINTS: PR00405; REVINTRACING.
DR PROSITE: PS00003; PH_DOMAIN; 3.
DR SMART: SM00105; ArfGAP; 1.
DR SMART: SM00233; PH; 3.
DR SMART: SM00324; RhoGAP; 1.
DR NON_TER 1
FT NON_TER
SQ SEQUENCE 1044 AA: 119453 MW: 68E22A6DEAA53BD CRC64;

Query Match 5.88: Score 103; DB 4: Length 1044;
Best Local Similarity 24.08; Pred. No. 2.3;
Matches 70; Conservative 39; Mismatches 119; Indels 64; Gaps 15;

QY 49 IVDOINVKLTCL---FSGNYEALPIIKNGDIVRFHRLKIQVYKKTQGITSSGFSALT 105
D 255 LLEETN-KKMCVLEGGFLSYENDKSTTPNGTININEVICIAHKEDFYLTNGPIF---I 310
QY 106 FEGTLGAPIIPRTSSKYFNFTIEDHKMVEALRVVA---STHMSPSNTLLKLCVQPMQYF 162
D 311 FE-----IYLPSEVFLFGAETS---QAQRKWTETAKHFEVPLFA-ENLEAD----Y 355
QY 163 DLTCOLLGKAEVDGASFLKLVW---DGTTRTPPSWRVLIDLVLEGLSHIHLQNLITID 219
D 355 DLICQLYK-DCHALDOWRKGNFAMDKSSLHF-----CLQMQEVOGDMHRLRLQELTIS 409
QY 220 ILVYDNH-VHVARSLKVGSLFLIYSLHTKL-----QSMNSNOT 257
D 410 IMVNGEKELDVLLIVEKGRITLYIHG-HTKLDFTVWHTAIEKAAAGTDGNALODQOLSKNDV 468
QY 258 MSLSEFLHFGTISYGRIPVLPESND-----VDOLKLDLESANLTANOH 302
D 459 PIIVNSCIAFTVQGLCKIYQKNGDPLHISELSEFFKDKDARSKFLRAGKH 520

```

```

RESULT 12
QY4E4 ID Q9H3V5 PRELIMINARY: PRI: 4370 AA.
AC Q9H3V5
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DI 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

```



```

|||||
Db 1211 CGCATCCGAGCAAAATTGAGGTCAATATAAGCCCAAGAGACTAATTCAGTCTGTAAACTT 1270
|||||
Qy 1261 cattgcccctaattgcttgcgaagaagattccaca 1298
|||||
Db 1271 CATTGCCCTAAATGTCATTGCTGCAAGAAGTTCACCA 1308
|||||

RESULT 3
AK022580 2383 bp mRNA PRI 29-SEP-2000
LOCUS Homo sapiens cDNA FLJ12518 fis, clone NT2RM2001805.
DEFINITION AK022580
ACCESSION AK022580
VERSION AK022580.1 GI:10434050
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA.
clone_lib:NT2RM2 clone:NT2RM2001805.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,
Matsumawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 2383)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (23-AUG-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- and 3'-end one pass sequencing and clone selection;
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
1..2383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone="NT2RM2001805"
/clone_lib="NT2RM2"
/notes="cloning vector: pME18SFL3-mRNA from uninduced NT2
neuronal precursor cells."
439..2343
/notes="unnamed protein product"
/codon_start=1
/protein_id="BABI4110.1"
/db_xref="GI:10434051"
/translation="MSLVPTNYYITPLNOLKGTIVNVGVKFFKPPVLSKGDYIC
SVTVIDQTVKPTCLLFSNGYEALPIIYKNGDILVRFHLKIQVKKETOGITSSGPA
SLTEFGTLKPIPTSSKYNFTEDHKWEALRVWASHMSPSWTLLKLDVOPNQ
YFDLTCQLLGAEDVASFLKVDWDTPTPPSWRVLQDLVLEGLDLSHLRLQNLTI
DILYVDHVAFLSKVSGFLRIYSLHKLOSMNSNOTMLSEFHLHGGSYGRGIR
VLPESSNDPOLKDLSEANLTANQSDVICQSEPDSDSFTSSGSVSLYEYERCOOLSA
TILTDHOLERTPLCALIKQKAPQOYIRAKLRYSKPRFLQSVKHLCPKCHLLQVYP
HEGLDIIIFDQATKTPDVKLONTSLYDSKIWTNKQGRKVAHVFKNNGILPLSNE
CLLIIEGTLSEICKLSNKTNSVIPVSGHEDLELLDLSAPFLIOGT IHHYGCKOCS
LRSIQNLNSLDYKSWIPSSVAEALGIYVQYVFMTFLDDGTGVLVLAIVMSDKFF
QIPASEVIMDDDLQKSDVMDIMDMFCPPGIKIDATPWLDECIFKSTNVINGINDQICVQI
FDTVAEDVI"
BASE COUNT 750 a 456 c 459 g 718 t
ORIGIN

Query Match 80.7%; Score 1048; DB 9; Length 2383;

```

```

Best Local Similarity 90.8%; Pred. No. 7.5e-268;
Matches 1178; Conservative 0; Mismatches 0; Indels 120; Gaps 1;

Qy 1 atgtcttgggttcagcaacaattatatataacacccctggaatcaacttaagggtggt 60
|||||
Db 439 ATGTCTTTGGTTCAGCAACAATAATATATATACACCCCTGAATCAACTTAAGGTTGT 498
|||||
Qy 61 acaattgtcaatgtctatggtgttggaagtctttaaagcccccatatcaagaaga 120
|||||
Db 499 ACAATTGTCAATGTCTATGTTGTGTGAAGTTCTTTAAGCCCCCATATCTFAAGCA 558
|||||
Qy 121 actgattattgtcagttgtaactattgtgaccagacaaatgtaaaactaaacttgctg 180
|||||
Db 559 ACTGATTATTGCTCAGTTGTAAGTATTGTGACCAGACAAATGTAAGTAACCTTGCTG 618
|||||
Qy 181 ctctttagtgaactatgtaagccctcccaataattataaaatggagatattgttcgc 240
|||||
Db 619 CTCTTTAGTGAACACTATGAGCCCTTCCAAATAATTATAAAATGGAGATATTGTTCG 678
|||||
Qy 241 ttccacaggtgaagattcaagtatataaaaggagactcaggtatcacagctctgc 300
|||||
Db 679 TTTACAGGCTGAAGATTCAAGTATATATAAAGAGAGACTCAGGGTATACACGCTCGG 738
|||||
Qy 301 ttgcatctttgacgtttgagggaactttggagccctcctatcatactcgcaacttcaag 360
|||||
Db 739 TTTGCATCTTTGACGTTTGAGGAACTTTGGAGCCCTATCATACCTCGCACTTCAAGC 798
|||||
Qy 361 aagtattttaactcactactgagaccacaaatgtagaagccttacgtgtttgggca 420
|||||
Db 799 AAGTATTTTAACTTCACTACTGAGGACCACAAATGTTAGAACCTTACGTGTGTTGG 858
|||||
Qy 421 tctactcatatgtcacctgtttggacattactaaaaattgtgtgatgttcagccaatgag 480
|||||
Db 859 TCTACTCATATGTCAACGCTCTTGGACATTAATAAATGTGTGATGTTTCAGCCAAATG 918
|||||
Qy 481 tatttgacgtcactgtcagctctggcgaagcagaagtgacgagcagcatcttctt 540
|||||
Db 919 TATTTTGACCTGACTGTTGACGCTTTGGCAAGCAAGTGGACGGAGCATCATTTCTT 978
|||||
Qy 541 cttaaagtgatgggagtcacacagacaccatttccattcttggagagtccttaatacaga 600
|||||
Db 979 CTAAAGGTATGGGATGGCAGCAGACACCACTTTCCATCTTTGGAGAGTCTTAATACA 1038
|||||
Qy 601 ctgttcttaagtgatttaagtccatccatcgcgcgtacacaaatctgacaatagacatt 660
|||||
Db 1039 CTGTGTTCTTGAAGTGATTTAAGTCACATCCATCGGCTACAAAATCTGACAAATAG 1098
|||||
Qy 661 ttagtctacgataaccatgttctatgtggcaagatctctgaaggttggaagctttcttag 720
|||||
Db 1099 TTAGTCTACGATAACCATGTTTCATGTGGCAAGATCTCTGAAGGTTTGAAGCTTTCTT 1158
|||||
Qy 721 atctatagccttcatacacaactcattcaatgaattcagagaatcagacaatgttaagt 780
|||||
Db 1159 ATCTATAGCCCTTCATACCAAACTTCAATCAATGAATTCAGAGAAATCAGACANTGTT 1218
|||||
Qy 781 ttagagtttctatcttcaggaggtaccagtttacggttcggggaatcagggttcgtccaga 840
|||||
Db 1219 TTAGAGTTTCTATCTTCATGAGGATACCAAGTTACGTCGGGGAATCAGGGTCTTGC 1278
|||||
Qy 841 agtaactctgatgtggatcaactaaaaagatttaataatcgaataattgacagcaaat 900
|||||
Db 1279 AGTAACCTCTGAIGTGGATCAACTGAAAAGGATTTAGAATCTGCAAAATTTGACAGCCA 1338
|||||
Qy 901 cagcattcagatgttatctgtcaatcagaacctgacagacagcttcccaaatggagctcg 960
|||||
Db 1339 CAGCATTACAGATGTTATCTGTCAATCAGAACCTCAGCAGACGCTTTCCAA----- 1387
|||||
Qy 961 ctctgctccagctggaggttcagtgagcagcggtctcggtctcattgagcgtccacctcc 1020
|||||
Db 1388 ----- 1387
|||||
Qy 1021 tgagttcaagcttctcctgcctcagctcccaagtagctgggattacaggctctggatca 1080
|||||

```


BASE COUNT 259 a 164 c 166 g 259 t 6 others
 ORIGIN

Query Match 27.4% Score 522.4; DB 10; Length 854;
 Host Local Similarity 99.3%; Pred. No. 20-115;
 Matches 545; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

1 atctcttcgctccagcaacaaatcatatcatcaacccctgaatccatctaaatgctg 60
 |||
 Db 132 atctcttcgctccagcaacaaatcatatcatcaacccctgaatccatctaaatgctg 191
 |||
 Oy 61 acacatctcaatgctcagctgctgctgaatctcttcaaccccatctcaatgctg 120
 |||
 Db 192 acacatctcaatgctcagctgctgctgaatctcttcaaccccatctcaatgctg 251
 |||
 Oy 121 acacatctcaatgctcagctgctgctgaatctcttcaaccccatctcaatgctg 180
 |||
 Db 252 acacatctcaatgctcagctgctgctgaatctcttcaaccccatctcaatgctg 311
 |||
 Oy 181 cctctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct 240
 |||
 Db 412 cctctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct 371
 |||
 Oy 241 cctctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct 300
 |||
 Db 372 cctctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct 431
 |||
 Oy 101 cctctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct 360
 |||
 Db 432 cctctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct 491
 |||
 Oy 361 aagctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct 420
 |||
 Db 492 aagctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct 551
 |||
 Oy 421 cctctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct 479
 |||
 Db 552 cctctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct 611
 |||
 Oy 480 cctctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct 538
 |||
 Db 612 cctctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct 671
 |||
 Oy 539 cctctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct
 |||
 Db 672 cctctctcagctcaatctcaacccctccatcaatctcaatctcaatctcaatct

RESULT 1:
 AV752694 529 bp mRNA EST 19-DEC-2000
 LACUS AV752694 NPD Homo sapiens cDNA clone NP24810 5', mRNA sequence.
 DEFINITION AV752694
 VERSION AV752694.1 GI:10910542
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 529)
 Song, H., Peng, Y., Gu, Y., Yang, Y., Guo, G., Xiao, H., Xu, X., Li, N.,
 Qian, B., Liu, F., Ou, J., Guo, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu,
 W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhou, M., Lu, G., Ye, M., Zhang, Q.,
 Han, Z., Chen, Z., Hu, R., and Chen, J.
 Homo sapiens NPD library cDNA clones
 UNPUBLISHED (2000)
 TITLE JOURNAL
 COMMENT Contact: Qianghua Zhang
 Shanghai Institute of Endocrinology, Rui-Jin Hospital
 197 Rui-Jin 11 Road, Shanghai 200025, P. R. China
 Tel: 86-21-64370045(ex. 663332)

FEATURES
 source
 1..529
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NP24810"
 /clone_lib="NPD"
 /tissue_type="Placental"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /note="Vector: pBluescript sk(-); site1: EcoRI; site2:
 XhoI"

BASE COUNT 168 a 106 c 104 g 150 t 1 others
 ORIGIN

Query Match 24.7% Score 469.6; DB 10; Length 529;
 Host Local Similarity 95.0%; Pred. No. 9-101;
 Matches 495; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

Oy 774 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 833
 |||
 Db 1 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 60
 |||
 Oy 834 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 893
 |||
 Db 61 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 120
 |||
 Oy 894 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 953
 |||
 Db 121 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 180
 |||
 Oy 954 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 1013
 |||
 Db 181 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 240
 |||
 Oy 1014 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 1073
 |||
 Db 241 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 300
 |||
 Oy 1074 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 1133
 |||
 Db 361 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 360
 |||
 Oy 1134 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 1193
 |||
 Db 361 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 419
 |||
 Oy 1194 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 1253
 |||
 Db 420 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 479
 |||
 Oy 1254 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 1294
 |||
 Db 480 gctaaagcttcaatctcaatctcaatctcaatctcaatctcaatctcaatctcaatct 520
 |||

RESULT 12
 AL577671 770 bp mRNA EST 16-FEB-2001
 LACUS AL577671 cDNA clone CS00K007Y014 3
 DEFINITION AL577671 cDNA clone CS00K007Y014 3
 VERSION AL577671
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

[illegible]

Bb	242	AGCTCCTACCGTTTCAAGCCGGCTTCCTCATFVACGAGGGAAGAAGGCATCATGTGATGT	301
OY	1510	aacagatgtctcgaatttcgaatacaaacaaacttaaatcccccgttgtaaagaatcg	1569
Dd	302	AACACTGCTGCAAGCTTGAAACCCTTAAAAATTAACTTAACCTCGCTTTTCATTAAGAACCG	761
OY	1570	tggatgctctctctcttctgcagaagaaacacagagatcatgatccccccaatacatgctttat	1629
Dd	162	TGGAGCTGCTCTCTCTCTGGAATAAAGACCTGTGGGCGCTGTGGCGCATGATGATGTTTTGTC	421
OY	1610	atcagctctcagccttcgaatacagaagaaagtactatgaagctctaccagaagatcttcac	1689
Dd	422	ATGCTTTTCATCTTGAGATGATATGGAAGAGCGAGCGTGTGGAAAGCTTACCTGAAAGATTTGAA	481
OY	1630	aaatctctccagctcttcagacacacaaaacctctcgaatgaatgaatgaatgaatgaatgaat	1749
Dd	482	CACCTTTTTCAGAGTCCGACAGCTGTAAGAGCTGCACATGATGACGAGCTTCAGACAGCTCA	541
OY	1750	gataatgacaatgaatcatgctcttctgctcccccgaagataaaaaatgtagatgcatctgag	1809
Dd	542	GAAACGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	601
OY	1810	gaatgctctcttcgaatgac	1868
Dd	602	GAGTGCCCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	661
OY	1869	gattcttgagac	1895
Dd	662	CA-TTTTGATFACATGACGTTGCCAGAAA	687
RESULT 15			
LOCUS	BC206213	546 bp	mRNA EST
DEFINITION	RST23-550 Albersys RAGE Library Homo sapiens cDNA, mRNA sequence.		21-APR-2001
ACCESSION	BC206213		
VERSION	BC206213.1 GI:13727900		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Pokratyuk; Melazon; Chordara; Granata; Vortelaput; Futeleostomi;		
AUTHORS	Mammaliat; Eubheria; Primatos; Carartini; Homunidae; Homo.		
	I (bases 1 to 546)		
	Harrington,J.,J., Short,B., Rundlett,S., Jackson,P.D., Perry,R.,		
	Gain,S., Dahl,T., Thornton,M., Ramchandran,K., Whittinton,J.,		
	Lerner,L., Krashinsky,D., McElinotte,K., Clark,S., Mays,R., Smith,E.,		
	Volsko,N., Hess,J., Colbreton,K., Lo,K., Oltendach,J., Danell,J.,		
	and Ducar,M.		
	Generation of Genome-wide Protein Expression Libraries using Random		
	Activation of Gene Expression		
JOURNAL	Nat. Biotechnol., 19 (3), 440 (2001) In press		
COMMENT	Contact: Scott J. Cain		
	Albersys, Inc.		
	1201 Carnegie Ave., Cleveland, OH 44115, USA		
	Tel.: 216 441 9900		
	Fax: 216 461 9596		
	Email: scott@albersys.com		
	High quality sequences. Smap: 511.		
FEATURES	Location/Qualifiers		
SOURCE	1..546		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/contig_lib="Albersys RAGE Library"		
	/cell_line="HT1080"		
	/note="See 'Generation of Genome-wide Protein Expression		
	Libraries using Random Activation of Gene Expression'		
	Nature Biotechnology, in press. Note that even though the		
	cell type indicated is HT1080, since a random activation		
	method was used, those sequence tags are not necessarily		
	expressed in HT1080 under normal circumstances."		
BASE COUNT	169 A 105 C 106 G 166 T		
ORIGIN			

```
Query Match      20.5%; Score 390.8; DB 11; Length 546;
Best local Similarity 99.5%; Pred. No. 9,5e-84;
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1506 atgtaaacagtgctctgagtttgaagatccatcaaaatctaaatccccctggttqatataaaac 1565
Db 1 atgtaaacagtgctctgagtttgaagatccatcaaaatctaaatccccctggttqatataaaac 60

Qy 1566 atcgtggattccctctctgaggaagagcactgggttatgtlaccctcccaataagtgatt 1625
Db 61 atcgtggattccctctctgaggaagagcactgggttatgtlaccctcccaataagtgatt 120

Qy 1626 tgttatgaccttiacacitgatgatggaacagggagtactagaagccctatctcatggattc 1685
Db 121 tctttatgacctttacacitgatgatggaacagggagtactagaagccctatctcatggattc 180

Qy 1686 tgacaatatcttccagatccagaculcagaagttctgagtgatgataaccttcaagaaga 1745
Db 181 tgacaatatcttccagatccagacatcagaagttctgagtgatgataaccttcaagaaga 240

Qy 1746 tctggatatgatcatggatatgtttgtccctccaggaataaaattgataatataccgtg 1805
Db 241 tctggatatgatcatggatatgtttgtccctccaggaataaaattgataatataccgtg 300

Qy 1806 attgaatgcttcatcaaatctatcacatgtcccaatgagacacataaatcaaatctacta 1865
Db 301 gttggaaatgcttcatcaatgctatcacatgtcccaatgagacacataaatcaaatctacta 360

Qy 1866 tcaaatlltttgaacacacagtttgcaggaagatata 1899
Db 361 tcagatTTTTGACACACACAGTTGCAGAGAGATCAA 394
```

Search completed: April 9, 2002, 22:13:52
Job time: 8867 sec

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 12, 2002, 08:33:16 : Search time 11.78 Seconds

(without alignments)
339,259 Million cell updates/sec

Title: US-09-816-248-5

Perfect score: 567

Sequence: 1 MSLVPATNYIVTLNQLKGG.....KKEIQGISSGFASLIFEGI 109

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs. 35664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	19.8	493	TEBA_STYMY	P29550 stylynychia
2	104.5	18.4	446	TEB_EUPCR	Q06184 euploies cr
3	97	17.1	495	TEBA_OXINO	P29549 oxytrichia n
4	90	15.9	450	TEBH_EUPCR	Q06183 euploies cr
5	72.5	12.8	1073	HSEB_PIG	P55204 sus scrofa
6	70.5	12.4	286	RIP1_CUCFI	Q9f1x4 cucumis fig
7	69.5	12.3	1161	DP3A_AQUAE	Q67125 aquifex aeo
8	68	12.0	1224	COPA_BOVIN	Q27854 bos taurus
9	67	11.8	87	SYR_SALTY	P74871 salmonella
10	67	11.8	1256	MEP_STRESU	P32653 streptococ
11	66.5	11.7	801	VP34_SCHPO	P50520 schizosacch
12	66	11.6	336	RUVB_HELPJ	Q9zm57 helicobacte
13	66	11.6	336	RUVB_HELPJ	Q25699 helicobacte
14	66	11.6	1251	RBP2_PLAVB	Q00799 plasmodium
15	65	11.5	221	GT24_ARATH	Q22vq3 arabidopsis
16	65	11.5	577	PRIM_BUCAI	P57184 buchnera ap
17	65	11.5	2244	PIYJ1_SCHPO	Q09794 schizosacch
18	64.5	11.4	412	FKB4_SPOFR	Q26486 spodoptera
19	64.5	11.4	577	SYR_HAEIN	P43832 haemophilus
20	63.5	11.2	122	Y012_NPVAC	P41422 autographa
21	63	11.1	433	SECY_RICPR	Q26c55 rickettsia
22	63	11.1	1224	COPA_HUMAN	P53621 homo sapien
23	62.5	11.0	1073	HSEB_HUMAN	Q36979 z genome po
24	62.5	11.0	3082	POLG_ZYMSV	P42001 caenorhabdi
25	62	10.9	282	YKCO_CAEEL	P25052 homo sapien
26	62	10.9	395	DDL_TREPA	Q83676 treponema p
27	62	10.9	421	LYPH_MYCGE	P47297 mycoplasma
28	62	10.9	624	YK08_YEAST	P54074 saccharomyc
29	62	10.9	877	DPO1_STRPN	P31252 streptococ
30	62	10.9	986	EP1B_STAEP	P30195 staphylococ
31	62	10.9	1450	MPSF_CHICK	Q02173 gallus gall
32	61.5	10.8	442	ETBR_PAT	P21451 rattus norv
33	61.5	10.8	928	NRPI1_XENLA	P28824 xenopus lae

34 61.5 10.8 1116 1 YK5A_AQUAE
35 61.5 10.8 2156 1 RRPL_PUMMH
36 61 10.8 338 1 SMIA_ASTBI
37 61 10.8 462 1 YB47_METJA
38 61 10.8 744 1 MCM5_HUMAN
39 60.5 10.7 251 1 TATC_RICPR
40 60.5 10.7 394 1 EX7L_THEMA
41 60.5 10.7 451 1 NRPL_ENTHR
42 60.5 10.7 569 1 TACT_HUMAN
43 60.5 10.7 633 1 C2AA_BACTK
44 60.5 10.7 656 1 FEMI_CAEEL
45 60.5 10.7 678 1 FLHA_AQUAE

ALIGNMENTS

```

RESULT 1
TEBA_STYMY STANDARD: PRT: 493 AA.
AC P29550:
DI 01-APR-1993 (Rel. 25, Created)
DI 01-APR-1993 (Rel. 25, Last sequence update)
DI 01-JUN-1994 (Rel. 29, Last annotation update)
DE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT.
GN STY56V AND STY561.
OS Stylynychia mytilis.
OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Stichotrichida;
OC Oxytrichidae; Stylynychia.
OX NCBI_TaxID:5952;
PN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92051290; Pubmed-1945829;
RA Fang G., Cech T.R.;
RI "Molecular cloning of telomere-binding protein genes from Stylynychia
  mytilis."
RT Nucleic Acids Res. 19:5515-5518(1991).
CC -!- FUNCTION: MAY FUNCTION AS PROTECTIVE CAPPING OF THE SINGLE-
CC -!- STRANDED TELOMERIC OVERHANG. MAY ALSO PARTICIPATE IN TELOMERE
CC -!- LENGTH REGULATION DURING DNA REPLICATION.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC -!- SUBCELLULAR LOCATION: NUCLEAR. STY56V IS SHOWN. STY561 DIFFERS IN
CC -!- MISCELLANEOUS: THE SEQUENCE OF STY56V BELONGS TO THE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT
CC -!- SIMILARITY: BELONGS TO THE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT
CC -!- FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X61749; CAA43888.1;
CC DR PIR: S22581; S22581.
CC InterPro: IPR003415; Telo_bind_alpha.
CC Pfam: PF02307; Telc_bind_alpha; 1.
CC DR DNA-binding; Nuclear protein; Telomere; Multigene family.
CC KW VZFIANT 429 429
CC FT VZFIANT 429 429
CC SEQUENCE 493 AA: 55v97 MW: 849002984363CSAA CRC64:

```

Query Match 11.8% Score 112; DR 1; Length 493;
Best Local Similarity 24.1% Pred. No. 9.2e-05;
Matches 30; Conservative 18; Mismatches 41; Indels 14; Gaps 2;
QY 5 ATRVITYPLNQ--LKGGTVNNVGVVKKFFPPYLSKGTDCSVTVIVD----- 51
:| | | | | :| | | | | :| | | | |
DB 30 STRKYVELNKASLTSAAQAHFYGVGVIDATFPYKTNQERYICSLKVVDPSPSLYLSKSGKTG 89
QY: 52 QTNVKLICLLFSGNVEALPIIYKNGDIVRFHRLKIQVKKETO 94


```
Db 115 RAGDIIRVHRATFLYNGORO 135
      : |||:| || :||:| :|
RESULT 4
TEBH_EUPCR STANDARD: PRT: 460 AA.
AC Q06183;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE TELOMERE-BINDING PROTEIN HOMOLOG.
OS Euplotides crassus.
OC Eukaryota: Alveolata: Ciliophora: hypotrichs: Euplotida: Euplotidae;
OC Euplotes.
OC NCBI_TaxID=5936;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=93126105; PubMed=1460483;
RA Wang W., Skopp R., Scofield M., Price C.;
PI telomere-binding protein homologs.;
PL Nucleic Acids Res. 20:5621-5629(1992).
CC -!- FUNCTION: MAY BIND TELOMERIC TAG4 SEQUENCES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M96819; AAA29128.1;
CC PIR: S35525;
CC InterPro: IPR003415; Telo_bind.alpha.
CC Pfam: PF02307; Telo_bind.alpha.1;
CC DNA-binding, Nuclear protein; Telomere; Multigene family.
CC KW SEQUENCE 460 AA: 53560 MW: EDIC141385A0B5FE CRC64;
-----
Query Match 15.9% Score 90: DB 1: Length 450;
Best Local Similarity 31.4% Pred. No. 0.022;
Matches 33: Conservative 12: Mismatches 46: Indels 14: Gaps 4:
Oy 9 YIYPLNOL--KGGIIVNYGVVKKFPPYLSKIDYCSVVIIVDQI-NVK-----LIC 59
Db 19 YEYTEIGSIEEENASINFYAVVIDACFPYKVDKKYKCYLKVIDITHNVKRGDDREAIIV 78
Oy 60 LLESGNVEALPIYKNGDIVRFHELKI-----QVYKKEITOGITSS 99
Db 79 ALQSRKFDLPIORCGDIIRVHRAEYNYKDDQHYFKLNYSYSS 123
-----
RESULT 5
HSEF_PIG STANDARD: PRT: 1073 AA.
AC P55204; Q29050;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HEAT-STABLE ENTEROTOXIN RECEPTOR PRECURSOR (GC-C) (INTESTINAL
DE GUANYLATE CYCLASE) (EC 4.5.1.2) (STA RECEPTOR).
GN GUCY2C OR GUC2C.
OS Sus scrofa (Pig).
OC Eukaryota: Metazoa: Chordata: Graniata: Vertebrata: Euteleostomi;
OC Mammalia: Eutheria: Cetartiodactyla: Suina: Suidae; Sus.
OX NCBI_TaxID=9823;
OX [1]
RN SEQUENCE FROM N.A.
RP
```

```
PC TISSUE=Small intestine;
EX MEDLINE=95058376; PubMed=7969686;
RA Wada A., Hirayama I., Kitao S., Fujisawa J.-I., Hidaka Y.,
RA Shimonishi Y.;
RI "pig intestinal membrane-bound receptor (guanylyl cyclase) for heat-
RI stable enterotoxin: cDNA cloning, functional expression, and
RI characterization.";
RL Microbiol. Immunol. 38:535-541(1994).
RN [2]
RP REVISIONS TO 238 AND 509.
RA Wada A.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: RECEPTOR FOR THE E-COLI HEAT-STABLE ENTEROTOXIN (E-COLI
CC ENTEROTOXIN MARKEDLY STIMULATES THE ACCUMULATION OF CGMP IN
CC HAMMALIAN CELLS EXPRESSING GC-C). ALSO ACTIVATED BY THE ENDOGENOUS
CC PEPTIDE GUANYLIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: GTP = 3', 5'-CYCLIC GMP + PYROPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE
CC FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PROTEIN KINASE-LIKE DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: D17513; BAA04455.1;
CC HSP: Q02846; IAWL.
CC InterPro: IPR001828; ANF_receptor.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR001054; Guanylyl_cyclase.
CC Pfam: PF01094; ANF_receptor; 1
CC Pfam: PF00211; guanylate_cyc; 1
CC Pfam: PF00069; pkinase; 1
CC SMART: SM00044; CYCC; 1.
CC PROSITE: PS00452; GUANYLATE_CYCLASES_1; 1.
CC PROSITE: PS00125; GUANYLATE_CYCLASES_2; 1.
CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC Receptor; Transmembrane; Glycoprotein; Lyase; CGMP synthesis; Signal.
KW SIGNAL.
FT CHAIN 1 24 1073 HEAT-STABLE ENTEROTOXIN RECEPTOR.
FT DOMAIN 24 433 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 434 454 POTENTIAL.
FT DOMAIN 455 1073 CYTOPLASMIC.
FT DOMAIN 489 749 PROTEIN KINASE-LIKE.
FT DOMAIN 824 954 GUANYLATE CYCLASE.
FT CATH 32 32 N-LINKED (GLCNAC...) (POTENTIAL).
FT CATH 43 43 N-LINKED (GLCNAC...) (POTENTIAL).
FT CATH 79 79 N-LINKED (GLCNAC...) (POTENTIAL).
FT CATH 195 195 N-LINKED (GLCNAC...) (POTENTIAL).
FT CATH 294 294 N-LINKED (GLCNAC...) (POTENTIAL).
FT CATH 307 307 N-LINKED (GLCNAC...) (POTENTIAL).
FT CATH 357 357 N-LINKED (GLCNAC...) (POTENTIAL).
FT CATH 402 402 N-LINKED (GLCNAC...) (POTENTIAL).
FT SEQUENCE 1073 AA: 123219 MW: 5644FC1327BA9F43 CRC64;
-----
Query Match 12.8% Score 72.5: DB 1: Length 1073;
Best Local Similarity 29.5% Pred. No. 4.7;
Matches 43: Conservative 19: Mismatches 39: Indels 21: Gaps 7:
Oy 14 LNOIKGGIIV---NHYGVVVFPPYLSKGT---DYCS---VVTIVDTIVKLTCLIFSG 64
Db 255 VHILKGGRAVAEDTIVILVDLFNDHYFMDNVTPADYKKNVLVLTPPENSVSNS--FSK 312
Oy 65 NWEAL----PIYKNGDIVRFHRLKIQVYKKEITOGITSSGFA----SLITEG 108
Db 312 ILSLVKNDIFLAYHNGVLLFQHMLKIFLEKRE--DVITSKFAHAFRNITEG 362
```

```

RESULT 6
ID RIPI_CUCFI STANDARD: PRI: 286 AA.
AC O9FRX4;
DI 20-AUG-2001 (Rel. 40, Created)
DI 20-AUG-2001 (Rel. 40, Last sequence update)
DI 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE RIBOSOME-INACTIVATING PROTEIN PRECURSOR (RRNA N-GLYCOSIDASE)
DE (EC 3.2.2.22).
OS Cucumis figareli.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eusoids I; Cucurbitales; Cucurbitaceae; Cucumis.
OX NCBI_TaxID=131071;
EN [1]
PP SEQUENCE FROM N.A.
RA Yamada T., Ohki S.T., Osaki T.;
RI "Cloning and analysis of a cDNA coding a putative ribosome-
PI inactivating protein from Cucumis figareli.";
PL Plant Biotechnol. 17:337-340(2000).
CC -!- CATALYTIC ACTIVITY: ENDOTHERMOLYSIS OF THE N-GLYCOSIDIC BOND AT ONE
CC -!- SPECIFIC ADENOSINE ON THE 28S RRNA.
CC -!- SIMILARITY: TO OTHER BACTERIAL AND PLANT RIBOSOME-INACTIVATING
CC PROTEINS. BELONGS TO TYPE 1 RIP.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB045560; BAB19677.1;
CC InterPro: IPR001574; RIP.
CC Pfam: PF00161; RIP; 1.
CC PRINTS: PR00396; SHIGARICIN.
CC PROSITE: PS00275; SHIGA_RICIN; 1.
KW Antiviral; Protein synthesis inhibitor; Hydrolase; Toxin; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 286 POTATIVE RIBOSOME-INACTIVATING PEPTIDE.
FT ACT_SITE 185 185 BY SIMILARITY.
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 286 AA; 31771 MW; 4EFD4966E604DA41 CRC64;
SQ
Query Match 12.4%; Score 70.5; DE 1; Length 286;
Best Local Similarity 25.6%; Pred. No. 1.7;
Matches 30; Conservative 13; Mismatches 37; Indels 37; Gaps 5;
QY 3 LVPATN-----YIVPLNOLKGGIIVNVYGVVKKFKPPYLSKGTIDYGSVVTIVDQ 59
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 63 LVPISGSPBYILMOLSNVEGRITIMADVIVNVYGVVKKFKPPYLSKGTIDYGSVVTIVDQ 117
QY 59 -----CLLFSONEALPIIYKNSDIVRHLKIQVKKKGTGTTSSGFAS; 104
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 118 SKVFOGKSIILPYSONYOKLO-----SVAPKEDSI-FLCFMFI 157
RESULT 7
ID DP3A_AQUAE STANDARD: PRI: 1161 AA.
AC O67125;
DI 15-DEC-1998 (Rel. 37, Created)
DI 15-DEC-1998 (Rel. 37, Last sequence update)
DI 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA POLYMERASE III ALPHA SUBUNIT (EC 2.7.7.7).
DN DNAA OR AQ.1009.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.

```

```

OX NCBI_TaxID=633563;
RH [1]
RF SEQUENCE FROM N.A.
RC STPAIN-VF5;
RX MEDLINE=98195656; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RI "The complete genome of the hyperthermophilic bacterium Aquifex
RI aeolicus.";
RI Nature 392:353-358(1998).
CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC N PYROPHOSPHATE + DNA(N).
CC -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC DIMERIZATION TO FORM THE POLIII' COMPLEX. POLIII' ASSOCIATES WITH
CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
CC AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-C FAMILY. DNAA
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000718; AAC07087.1;
CC InterPro: IPR003141; PHE-N.
CC DR InterPro: IPR002309; tRNA-synt_2.
CC Pfam: PF02231; PHE-N; 1.
CC DR Pfam: PF01335; tRNA-anti; 1.
CC DR SMART: SM00481; POLIIIAC; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW Complete proteome.
SQ SEQUENCE 1161 AA; 133207 MW; 619F7436E1262BD4 CRC64;
Query Match 12.3%; Score 69.5; DB 1; Length 1161;
Best Local Similarity 29.3%; Pred. No. 11;
Matches 22; Conservative 12; Mismatches 24; Indels 17; Gaps 4;
QY 5 PATNY-----IYPLNLK-----GGTIVNVYGVVKKFKPPYLSKGTIDYGSVVTIVDQ 52
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 972 PLDNVEKLLKNRYTPIEDLEENDKESEAVLTGVITELKVKK----TKNGDYMAVENLVDK 1027
QY 53 TNVKTCLLESQNE 67
||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 1028 TGL-IECVFPGVYE 1041
RESULT 8
ID CCPA_BOVIN STANDARD: PRI: 1224 AA.
AC Q27554;
DI 01-NOV-1997 (Rel. 35, Created)
DI 01-NOV-1997 (Rel. 35, Last sequence update)
DI 20-AUG-2001 (Rel. 40, Last annotation update)
DE CALCIMER ALPHA SUBUNIT (ALPHA-COAT PROTEIN) (ALPHA-COP) (HEPCOP)
DE (HEP-COP) (CONTAINS: XENIN (XENOPSIN-RELATED PEPTIDE); PROXENIN).
GH CCPA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;

```



```

RI Streptococcus suis type 2.*;
RL Infect. Immun. 60:2351-2367(1992).
CC -!- SUBCELLULAR LOCATION: TYPE 1 MEMBRANE PROTEIN. CELI WALL.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PEPTIDINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X64450; CAA45781.1;
DR PIR: A43829; A43822.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00745; Gram_pos_anchor_1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KW Repeat: Transmembrane; Cell wall; Signal.
FT SIGNAL 1 47
FT CHAIN 48 1256 MURAMIDASE-RELEASED PROTEIN.
FT DOMAIN 48 1231 EXTRACELLULAR (POTENTIAL).
FT TRANSHEM 1232 1248 MEMBRANE ANCHOR.
FT DOMAIN 1249 1256 CYTOPLASMIC (POTENTIAL).
FT REPEAT 563 681 SMALL.
FT REPEAT 839 851 SMALL.
FT REPEAT 953 1005 LAPCE.
FT REPEAT 1064 1084 SMALL.
FT REPEAT 1089 1142 LARGE.
FT REPEAT 1143 1195 LARGE.
FT DOMAIN 1223 1228 CONSERVED IN GRAM-POSITIVE CELL SURFACE
SO SEQUENCE 1256 AA; 135795 MW; DCF7F5242F14341 CRC64;

Query Match 11.8%; Score 67; DB 1; Length 1256;
Best Local Similarity 29.2%; Pred. No. 22;
Matches 28; Conservative 15; Mismatches 33; Indels 2; Gaps 5;

OY 7 TNYIYPLNOLKGTIVNVY-----GVVKKFKP----PYLSKGTIDYCVTVIYDCTTKL 57
DB 1000 TTHYKTPKNGKIVNVNVTEDGVTK--EPVDTFTSPGTPY-----DILCKRP 1050
OY 58 TLLFSG-NYEALPII-YKNGDIVRHLKIOVYKK 91
DB 1051 KIITFKGEYELVVDGIENCKVVEGETIVTVYVYRK 1086

RESULT 11
VP34_SCHPO
ID VP34_SCHPO STANDARD: PPT: 801 AA.
AC P50520; Q3P3W3; Q9UPD2;
DI 01-OCT-1996 (Pel. 34, Created)
DI 20-AUG-2001 (Pel. 40, Last sequence update)
DI 20-AUG-2001 (Pel. 40, Last annotation update)
DE PHOSPHATIDYLINOSITOL 3-KINASE VPS34 (EC 2.7.1.137) (PI3-KINASE)
DE (PTDINS-3-KINASE) (PI3K) (VACUOLAR SORTING PROTEIN 34).
GN VPS34 OR SPAC458.05.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID:4896;
[1]
PN
PP SEQUENCE FROM N.A.
EX MEDLINE=96350252; Pubmed=8719881;
PA Ikegawa K., Dewald D.B., Emr S.E.;
FT Schizosaccharomycetes pombe vps34p, a phosphatidylinositol-specific PI
FT 3-kinase essential for normal cell growth and vacuole morphology.*;
PL J. Cell Sci. 106:3745-3756(1995).
[2]
PP SEQUENCE FROM N.A.

```

```

PC STEIN-972;
PA Harris D., Lyne M., Rajandream M.A., Barrell B.G.;
RN Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
[3]
PP SEQUENCE OF 138-801 FROM N.A.
EX MEDLINE=95290763; Pubmed=7772832;
PA Kimura K., Miyake S., Makuuchi M., Morita R., Usui T., Yoshida M.,
PA Harinouchi S., Fukui Y.;
FT Phosphatidylinositol 3-kinase in fission yeast: a possible role in
FT stress responses.*;
RL Biosci. Biotechnol. Biochem. 59:678-682(1995).
CC -!- FUNCTION: PHOSPHATIDYLINOSITOL 3-KINASE HOMOLOG REQUIRED FOR
CC VACUOLAR SORTING AND SEGREGATION.
CC -!- CATALYTIC ACTIVITY: ATP + 1-PHOSPHATIDYL-ID-MYO-INOSITOL = ADP +
CC 1-PHOSPHATIDYL-ID-MYO-INOSITOL 3-PHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE PI3/P14-KINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U92593; AAC19133.1;
DR EMBL: AL358272; CAB93947.1;
DR InterPro: IPR002420; PI3K_C2.
DR InterPro: IPR001263; PI3Ka.
DR InterPro: IPR000403; PI3_P14_kinase.
DR Pfam: PF00613; PI3Ka; 1.
DR Pfam: PF00792; PI3K_C2; 1.
DR Pfam: PF0454; PI3_P14_kinase; 1.
DR SMART: SM00145; PI3Ka; 1.
DR SMART: SM00146; PI3K; 1.
DR SMART: SM00142; PI3K_C2; 1.
DR PROSITE: PS00915; PI3_4_KINASE_1; 1.
DR PROSITE: PS00916; PI3_4_KINASE_2; 1.
DR PROSITE: PS00290; PI3_4_KINASE_3; 1.
KW Transferase; Kinase.
FT DOMAIN 541 799 PI3K/PI4K.
FT CONFLICT 164 164 K -> O (IN REF. 1).
FT CONFLICT 236 236 L -> I (IN REF. 3).
FT CONFLICT 542 542 S -> T (IN REF. 3).
SO SEQUENCE 801 AA; 92135 MW; 520571E1475CC341 CRC64;

Query Match 11.7%; Score 66.5; DB 1; Length 801;
Best Local Similarity 29.9%; Pred. No. 15;
Matches 3; Conservative 9; Mismatches 24; Indels 21; Gaps 3;

OY 23 VVVYGVVKKFKPFLVSKGIDYCVTVIYDCTTKLFLSGNVEALPIIYKNGDIVRPH 82
DB 509 VRIYVRII-----PDAC---TVFKSTMQLRLFLFKCQDGSKYPIIIFKNGDRLQRD 554
OY 93 RIKYIV-----YKKE 92
DB 555 CVVQQLILMDKLIKKE 571

RESULT 12
RUVF_HELPJ
ID RUVF_HELPJ STANDARD: PPT: 336 AA.
AC Q92M57;
DI 30-MAY-2000 (Pel. 39, Created)
DI 30-MAY-2000 (Pel. 39, Last sequence update)
DI 20-AUG-2001 (Pel. 40, Last annotation update)
DE HOLLIVET JUNCTION RNA HELICASE RUVF.
GN RUVF - P. JPH0365.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID:85963;

```

[1] SEQUENCE FROM N.A.
PP MEDLINE=99120557; PubMed=9923682;
RX Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Bois P.C.,
PA Smith D.P., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
PA Gibson R., Merberg D., Mills S.D., Jiang O., Taylor D.E., Voris G.F.,
RA Trust J.J.;
FA "Genomic sequence comparison of two unrelated isolates of the human
RI gastric pathogen Helicobacter pylori";
RL Nature 397:176-180(1999).
CC -!- FUNCTION: THE RUVA-RUVB COMPLEX IN THE PRESENCE OF ATP REPAIRS
CC CRUCIFORM STRUCTURE IN SUPERCOILED DNA WITH PALINDROMIC SEQUENCES,
CC INDICATING THAT IT MAY PROMOTE STRAND EXCHANGE REACTIONS IN
CC HOMOLOGOUS RECOMBINATION. RUVA IS AN HELICASE THAT MEDIATES THE
CC HOLLIDAY JUNCTION MIGRATION BY LOCALIZED DENATURATION AND
CC REANNEALING (BY SIMILARITY).
CC -!- SUBUNIT: FORMS A COMPLEX WITH RUVA (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE RUVB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in "any way"
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/about/license/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE001471; AAC05938.1; -.
DR InterPro: IPF003593; AAA.
DR InterPro: IPF001939; AAA-subfam.
DR Pfam: PF00004; AAA; 1.
DR SMART: SM00382; AAA; 1.
KW DNA repair; SOS response; ATP-binding; DNA recombination; Helicase.
KW Complete proteome. 67 ATP (POTENTIAL).
KW NP_BIND 50 37280 MW: 4096BIAI8CA4416 CRC64:
SQ SEQUENCE 336 AA: 37280 MW: 4096BIAI8CA4416 CRC64:

Query Match 11.5% Score 66; DB 1; Length 336;
Best Local Similarity 35.4%; Pred. No. 6.4;
Matches 17; Conservative 8; Mismatches 13; Indels 10; Gaps 2;

Oy 31 FFKPPYKSGTDYCSVTIVDTNWKLTCLLFSGHVEALPIYKNEDI 78
||||| : : ||||| 1 1 1 1 1 1
Db 58 FGPGGLCK-ISIHIIAKEMETIKIT-----AAPHERSGDL 95

RESULT 13
FOVB_HELPY STANDARD: PRT: 336 AA.
AC FUVB_HELPY
ID Q25699;
DI 15-JUL-1998 (Rel. 35, Created)
DI 15-JUL-1998 (Rel. 36, Last sequence update)
DI 20-AUG-2001 (Rel. 40, Last annotation update)
DE HOLLIDAY JUNCTION DNA HELICASE RUVB.
GN RUVB OR HP1059.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group.
OC Helicobacter.
OX NCBI_TaxID=210;
PN [1]
PP SEQUENCE FROM N.A.
PC STRAIN=26595 / ATCC 700392;
RX MEDLINE=97394457; PubMed=9252185;
PA Tomb J.-F., White O., Kervatage A.P., Clayton P.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
PA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glöckle A.K.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey F.K.,
RA Berg D.E., Gosyne J.D., Utterback T.R., Peterson J.D., Keller J.M.,
PA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin P.,
PA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.

```

PA Venter J.C.:
CC -The complete genome sequence of the gastric pathogen Helicobacter
CC pylori.
CC FL Nature 388:530-547(1997).
CC CC
CC -f- FUNCTION: THE RUVA-RUVB COMPLEX IN THE PRESENCE OF ATP RENATURES
CC CRISPR/CMR STRUCTURE IN SUPERCOILED DNA WITH PALINDROMIC SEQUENCE.
CC INDICATING THAT IT MAY PROMOTE STRAND EXCHANGE REACTIONS IN
CC HOMOLOGOUS RECOMBINATION. RUVAB IS AN HELICASE THAT MEDIATES THE
CC HOLLIDAY JUNCTION MIGRATION BY LOCALIZED DENATURATION AND
CC REANNEALING (BY SIMILARITY).
CC -f- SUBUNIT: FORMS A COMPLEX WITH RUVA (BY SIMILARITY).
CC -f- SIMILARITY: BELONGS TO THE RUVB FAMILY.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce,
CC or send an email to license@isb-sib.ch).
CC CC
CC EMBL: AF00613; AM008100.1;
CC FL TIGR: AF1059;
CC DR InterPro: IPR001593; AAA_subfam.
CC DR InterPro: IPR001593; AAA_subfam.
CC DR Pfam: PF000004; AAA: 1.
CC DR SMART: SM00362; AAA: 1.
CC DR DNA repair: SOS response; ATP-binding; DNA recombination; Helicase;
CC KW Complete proteome.
CC FL NP_019816 50
CC SQ SEQUENCE 335 AA: 37360 MW: 4FA54074AF6F5A0 CWC64;
CC CC
CC Query Match 11.6% Score 56; DB 1; Length 336;
CC Best Local Similarity 35.4%; Pred. Mismatches 13; Indels 10; Gaps 2;
CC Matches 17; Conservative 8;
CC CC
CC Cy 31 FFKPPYLSKTDYCSVTVTDQTVKLTCLFSGNVEALPIYKNGDI 78
CC ||||| : : |||:| | :|:|:|
CC Tb 58 FFGPPGLCK-TSISHIAKEHETNIKIT-----AAPMIKSGDL 95
CC CC
CC RESULT 14
CC RBP2_PLAYB STANDARD: PRT: 1251 AA.
CC AC Q00799;
CC DT 01-APR-1993 (Rel. 25, Created)
CC DT 01-APR-1993 (Rel. 25, Last sequence update)
CC DI 01-OCT-1995 (Rel. 34, Last annotation update)
CC DE PETICULOCYTE BINDING PROTEIN 2 (FRAGMENT).
CC GN RBP2
CC OS Plasmodium vivax (strain Relem).
CC CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CC CC NCBI_Taxid:31273;
CC PR |||
CC PP SEQUENCE FROM N.A.
CC PX MEDLINE:52315343; PubMed:1617731;
CC PA Galinski M.P., Medina C.C., Inoravalle F., Barnwell J.W.;
CC ET "A reticulocyte-binding protein complex of Plasmodium vivax
CC parasites".
CC PL Cell 53:1213-1226(1992).
CC CC -f- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -f- SUBCELLULAR LOCATION: MEMBRANE-BOUND (PROBABLE).
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).

```

QY 8 NYIYPLNOLKGGTI-----VNVGVV-----KFKPPYLSK 39
Db 33 DYIYIPVNLKGGDOFSDFKKINPMGTVPALVGDVWVINDSFALLMYLDEKYPPEPLLP 92
QY 40 GTDYCSVVTIYDQTNVKTCLLFSNGYEALPII-----YKNGDIVRFHRLKIQVYRKE- 92
Db 93 DLHKEAV-----NYQMSIVLSGIOPHONLAVIRYIEKINVEEKTA 134
QY 93 -TCGITSSEFASL 104
Db 135 WYNNAITKGFTAL 147

Search completed: April 12, 2002, 08:33:38
Job time: 22 sec

DE EMBL: M88098: AAA29744.1: -
KW Malaria: Peceptor; Membrane.
FI NON_TER 1251 1251
FT NON_TER 1251 1251
SQ SEQUENCE 1251 AA: 143741 MW: 54BA51C7404AC572 CRC64:

Query Match 11.6% Score 66: DB 1: Length 1251:
Best Local Similarity 26.3%: Pred. No. 26:
Matches 26: Conservative 11: Mismatches 28: Indels 34: Gaps 4:

QY 8 NYIYPLNOLKGG--GTIYVNVGVVFKFPYLSKGTIDYCSVVTIYDQTNVKTCLLFSNG 55
Db 492 NLIVTEKNELNGIDSIINIEGALKESK-----GN 521
QY 55 YEALPIYKNCDIVFPHRLKIQVYKKEIOGITSSEFASL 104
Db 522 YE-IGFLEKEEIGCKNRKLVDTIRKSINS-TVGNFSSL 558

RES'LI 15
GI72_APATH STANDARD: PRT: 221 AA.
AC Q92VQ3:
DI 20-AUG-2001 (Rel. 40, Created)
DI 20-AUG-2001 (Rel. 40, Last sequence update)
DI 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE GLUTATHIONE S-TRANSFERASE ZETA-CLASS 2 (EC 2.5.1.18).
GN AT2G02390 OR T16F16.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-J., Joun C.D.,
RA Fujii C.Y., Mason I.M., Bowman C.L., Barnstead M.E., Feldblum I.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen J.E., VanAken S.E., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy I.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RI "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana";
PL Nature 402:761-768(1999).
CC -!- CATALYTIC ACTIVITY: PX + GLUTATHIONE = HX + R-S-G.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. ZETA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AC005312; AAC78521.1;
CC MF000521; GST.
CC InterPro: IPR000521; GST.
CC Pfam: PF00043; GST; 1.
CC Transferase.
FT ACT_SITE 17 17 POTENTIAL.
SQ SEQUENCE 221 AA: 24897 MW: 106BDC8EF3E745BF CRC61:

Query Match 11.5% Score 65: DB 1: Length 221:
Best Local Similarity 21.1%: Pred. No. 5.1:
Matches 28: Conservative 18: Mismatches 33: Indels 54: Gaps 5:

QY 8 NYIYPLNOLKGG--GTIYVNVGVVFKFPYLSKGTIDYCSVVTIYDQTNVKTCLLFSNG 55
Db 492 NLIVTEKNELNGIDSIINIEGALKESK-----GN 521
QY 55 YEALPIYKNCDIVFPHRLKIQVYKKEIOGITSSEFASL 104
Db 522 YE-IGFLEKEEIGCKNRKLVDTIRKSINS-TVGNFSSL 558

A: Introns: 32/3
C: Keywords: DNA binding; nucleus

Query Match 19.8% Score 112: DB 2: Length 493;
Best Local Similarity 29.1% Pred. No. 0.00033;
Matches 30: Conservative 18: Mismatches 41: Indels 14: Gaps 2:
QY 6 ATNYITPLNO--LGGTIVNVYGVVKKFPYLSKGTIDYCSVVTIYD----- 51
: | | | | | : | | | | | : | | : | |
Db 30 STRKYVELNKASLTSAEAQHFYGVDAITFPYKINGERYICSLKVVDPDPSLYLKSCKGTG 89
: | | | | | : | | | | | : | | : | |
QY 52 OTNVKLTCLIFSGNYEALPIIYKNGDIVRFHRLKIQVYKKEIO 94
: | | | | | : | | | | | : | | : | |
Db 90 DASOYAILVLAQKFDPLPIHRIIGDIIRVHRATILYNGQRQ 132

RESULT 3
S35524
telomere-binding protein, 51K - Euplotes crassus
C: Species: Euplotes crassus
C: Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-Dec-1999
C: Accession: S35524
R: Wang, W.; Skopp, R.; Scofield, M.; Price, C.
Nucleic Acids Res. 20, 6621-6629, 1992
A: Title: Euplotes crassus has genes encoding telomere-binding proteins and telomere-binding proteins
A: Reference number: S35524; MUID: 93126105
A: Accession: S35524
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-445 <WAN>
A: Cross-references: EMBL: M96818; NID: g159023; PID: g159024
C: Genetics:
A: Genetic code: SGC9
A: Introns: 9/3
C: Keywords: DNA binding; nucleus

Query Match 18.4% Score 104.5: DB 2: Length 445;
Best Local Similarity 33.7% Pred. No. 0.0019;
Matches 33: Conservative 14: Mismatches 38: Indels 13: Gaps 4:
QY 8 NYITPLNQLK--GGTIVNVYGVVKKFPYLSKGTIDYCSVVTIYDQTNV-----K 54
: | | | | | : | | | | | : | | : | |
Db 11 HVOYSDLSIKKEEEDQYHFYGVVIDASFPY--KGEKRVVTKVADPSSVAKGRLNT 68
: | | | | | : | | | | | : | | : | |
QY 57 LTCLIFSGNYEALPIIYKNGDIVRFHRLKIQVYKKEIO 94
: | | | | | : | | | | | : | | : | |
Db 59 VNVVFFSONFEDFLIIOFVGDIVPVHRAQLQHYNDKQ 105

RESULT 4
A41221
macronuclear alpha protein (alanine version) - Oxytricha nova
C: Species: Oxytricha nova
C: Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 02-Mar-2001
C: Accession: A41221; B41221
R: Gray, J.T.; Celandier, D.W.; Price, C.M.; Cech, T.R.
Cell 67, 807-814, 1991
A: Title: Cloning and expression of genes for the Oxytricha telomere-binding protein: sp
A: Reference number: A41221; MUID: 92035001
A: Accession: A41221
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-495 <GRA>
A: Cross-references: GB: M58931; NID: g159811; PID: g159812
A: Accession: B41221
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-310, 'S', 312-455, 'E', 457-495 <GP2>
A: Cross-references: GB: M68230; NID: g159813; PID: g159814
C: Genetics:
A: Genetic code: SGC5

C: Keywords: DNA binding; nucleus

Query Match 17.1% Score 97: DB 2: Length 495;
Best Local Similarity 30.9% Pred. No. 0.014;
Matches 29: Conservative 14: Mismatches 30: Indels 12: Gaps 1:
QY 25 YGVVKKFPYLSKGTIDYCSVVTIYDQ-----NVKLTCLIFSGNYEALPIIY 73
: | | | | | : | | | | | : | | : | |
Db 55 YAVVIDATFPYKINGERYICSLKIVDPTLYLKQOKGAGDASVATLVYAKREDPLIIH 114
: | | | | | : | | | | | : | | : | |
QY 74 KNGDIVRFHRLKIQVYKKEIO 94
: | | | | | : | | | | | : | | : | |
Db 115 RAGDIIRVHRATILYNGQRQ 135

RESULT 5
S35525
telomere-binding protein - Euplotes crassus
C: Species: Euplotes crassus
C: Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-Dec-1999
C: Accession: S35525
R: Wang, W.; Skopp, R.; Scofield, M.; Price, C.
Nucleic Acids Res. 20, 6621-6629, 1992
A: Title: Euplotes crassus has genes encoding telomere-binding proteins and telomere-binding proteins
A: Reference number: S35524; MUID: 93126105
A: Accession: S35525
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-460 <WAN>
A: Cross-references: EMBL: M96819; NID: g159025; PID: g159025
C: Genetics:
A: Genetic code: SGC9
A: Introns: 18/3; 377/3
C: Keywords: DNA binding; nucleus

Query Match 15.9% Score 90: DB 2: Length 460;
Best Local Similarity 31.4% Pred. No. 0.071;
Matches 33: Conservative 12: Mismatches 46: Indels 14: Gaps 4:
QY 9 YIYTPLNQL--KGTIVNVYGVVKKFPYLSKGTIDYCSVVTIYDQTNV-----LTC 59
: | | | | | : | | | | | : | | : | |
Db 19 YEYETIGSTEEENEASINFYAVVIDACFPYKVDKKYKLVKVIDITHNVKGGDNDFAIV 78
: | | | | | : | | | | | : | | : | |
QY 60 LIFSCHYEALPIIYKNGDIVRFHRLKI-----QVYKKEIOGITSS 99
: | | | | | : | | | | | : | | : | |
Db 79 ALQSRKEEDPLIIOFGDIIRVHRAEYNYKDDQHYFKLNMSYSS 123

RESULT 6
A84466
hypothetical protein At2g05210 [imported] - Arabidopsis thaliana
C: Species: Arabidopsis thaliana (mouse-ear cress)
C: Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C: Accession: A84466
R: Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Renito, M.L.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon
Buss, D.; Nicornan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ventre
Nature 402, 761-768, 1999
A: Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A: Reference number: A84466; MUID: 20083487
A: Accession: A84466
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-318 <STO>
A: Cross-references: GB: AE002093; NID: g4755192; PID: AAD29059.1; GSFDB: GN00134
C: Genetics:
A: Gene: At2g05210
A: Map position: 2

Query Match 15.0% Score 85: DB 2: Length 318;

Best Local Similarity 30.4%; Pred. No. 0.16;
Matches 21; Conservative 16; Mismatches 30; Indels 2; Gaps 1;

Oy 23 VNYGVVWFPPYLLSKGIDYCSVVTVVDOT--NVKLTCLLFSGNVEALPIYKNGEDIVR 80
I:: I::
DB 24 VSLGIQVLEOPEKOCRRNDWDICTIRIIDDTYPSPGLIVNWFSKITLEOLPQIKNHUHL 93
I:: I:: :

Oy 81 FHRPKIQVY 89
I:: I::
DB 84 FTRIKWQIF 92
I:: I:: :

RESULT 7

B86210
protein F22G5.6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B86210
A:Authors: Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Crossley, T.H.; Dewar, R.
Nansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, e15-820, 2000
A:A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maithi, P.; Marziani,
Rizzo, M.; Rooney, T.; Salter, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B86210
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-937 <STO>
A:A:Cross-references: GB:AE005172; MID:q8778570; PIDN:AAF79578.1; GSTEB:GN00141
C:Genetics:
A:Gene: F22G5.6
A:Map position: 1

Query Match 12.8%; Score 72.5; DB 2; Length 937;
Best Local similarity 24.6%; Pred. No. 13;
Matches 29; Conservative 15; Mismatches 33; Indels 41; Gaps 5;

Oy 3 LVPAINVI-----YT-PLNOL-----KGGTIVNWVGWVFKEP--- 34
I:: I::
DB 93 LLPLSGFISDCGLQADESYTEPLTKLFITSDADFKSKGKIQNVGP-MEVIKVIV 151
I:: I:: :

Oy 35 -PYLSKGTDYCSVVTVVDOTNVKLTCLLFSGNVEAL-----PIYKNGEDIVR 80
I:: I::
DB 152 LVFPDGVPCVTILIVIOGINVLIVAMFTVGYNDLTHHPFDLYLGPNHWIVDI..F 209
I:: I:: :

RESULT 9

A93897
hypothetical protein BH1977 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: A93897
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai,
Nucleic Acids Res. 28, 4317-4331, 2000
A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20253314
A:Accession: A83897
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1054 <STO>
A:Cross-references: GB:AF001513; GB:BA000004; MID:g10174345; PIDN:BAE05696.1; GSPDB:GSPDB:
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1977

Query Match 12.6%; Score 71.5; DB 2; Length 1054;

A:Cross-references: GB:299110; GB:AL009126; NID:g2633472; PIDN:CAB13067.1; PID:ell832303
A:Experimental source: strain 158
C:Genetics:
A:Gene: YjeA
C:Superfamily: nodB homology
F:283-428/Domain: nodB homology <NODE>

Query Match 12.3% Score 69.5; DB 2; Length 467;
Best Local Similarity 28.0%; Pred. No. 12;
Matches 23; Conservative 16; Mismatches 36; Indels 7; Gaps 2;
OY 22 IVNVYGVKFEKPPYLSKGDYCSWTVTDQINVKLTCLLFSG--NYEALPIIYKH---75
DB 15 VVCVGLIGFHHSLKKEIVMKVVRTDSOYGVETAILVNDGKTFYAYNPVFKNEKM 74
OY 76 -GDIVFHLRLKIOYVKETQCI 96
DB 75 DSALKREAEKEVCFQKTRDV 96

RESULT 11
C84517
probable receptor-like protein kinase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84517
R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umarani, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 751-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: C84517
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-885 <STO>
A:Cross-references: GB:AE002093; NID:g4263827; PIDN:AAD15470.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g14440
A:Map position: 2

Query Match 12.3% Score 69.5; DB 2; Length 886;
Best Local Similarity 23.5%; Pred. No. 25;
Matches 23; Conservative 17; Mismatches 29; Indels 29; Gaps 4;
OY 5 PATNVIV-IPNLNKGSHIYV-----YGVKFEKPPYLSKGDYCSWTVTD 51
DB 46 PLIMLTISDVNVPGCKTNKKNHSDITSPPYKVLRIF-----PEGIFNYSLSVKO 100
OY 52 OTNRYKLCILFSGNYEAL-----PIIYKHGDI 78
DB 101 GTKYLIRTLFFYGNVDGLNTSPREDLFGPNQIWTSVDV 138

RESULT 12
B70387
DNA polymerase III alpha subunit - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 18-Jun-1999
C:Accession: B70387
R:Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O.
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196566
A:Accession: B70387
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1151 <AQF>
A:Cross-references: GB:AE000718; NID:g2983504; PIDN:AA007087.1; PID:g2983510; GB:AE000696
A:Experimental source: strain VF5

```

C:Genetics:
A:Gene: dnaE
C:Superfamily: DNA-directed DNA polymerase III alpha chain

Query Match      12.3%   Score 69.5;   DB 2;   Length 1161;
Best Local Similarity 29.3%;   Pred. No. 34;
Matches 22;   Conservative 12;   Mismatches 24;   Indels 17;   Gaps 4;

Oy  5  PATHY-----IYFLNQLK-----GGIIVNVGVVKKFKPPYLSKGTDCSVTIDQ 52
      |||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  972 PLDNYELKLRNPTIEDLEWDEKSEAVLTGVITELKVKK---TKNGDYMAVENLVDK 1027
      |||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Oy  53  TNVLCILFSGNVE 57
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  1028 TGLTECVWFFGVYE 1041
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

RESULT 13
E75338
DNA gyrase, subunit A - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: E75338
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.C.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: E75338
A>Status: preliminary
A:Molecule type: DNA
A:Cross-references: GB:AE002030; GB:AE000513; MID:g6459692; FIDN:ANF11467.1; PID:g645
A:Experimental source: Strain R1
C:Genetics:
A:Gene: DR1913
A:Map position: 1
C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A: phage T4 DNA topoisomerases

Query Match      12.1%   Score 68.5;   DR 2;   Length 812;
Best Local Similarity 27.0%;   Pred. No. 29;
Matches 27;   Conservative 21;   Mismatches 39;   Indels 13;   Gaps 5;

Oy  1  MSLVPAINIYITPLNQLKGGTI-VNVGVVKKFK-PPYLSKGTDCSVWT--IVDOIIVK 56
      |||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  599 MALVPG-----NDEGELLALISEGLKRTKVSVDYPSKRGGLGVITLDVTDKTKL 750
      |||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Oy  57  LTCLFFSGNYEALPIIYKNGDVIHFRLKIQVYKKTGGI 96
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db  751 VTLAHVAGN-EELMLVTEKGVITRTVEEVRVTRGNAQGV 789
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

RESULT 14
B83723
Hypothetical protein BH0546 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: B83723
R:Ikami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans A
A:Reference number: B83650; MUID:20263314
A:Accession: B83723
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1,306 <STC>
A:Cross-references: GB:AF001509; GB:BA0000004; MID:g10173176; PIDN:BAR04305.1; GSFD8:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0546

```

Search completed: April 12, 2002, 08:33:59
Job time: 43 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

QW protein - protein search, using sw model

Run on: April 12, 2002, 08:33:16 ; Search time 23.56 seconds
(without alignments)
341.257 Million cell updates/sec

Title: US-09-816-248-5
Perfect score: 567
Sequence: 1 MSIVPATNYVTPLNOKGG.....KKETQITSSGCAFLIFET 109

scoring table: BLOSUM52
Gapop 10.0 , Gapext 0.5

searched: 522463 segs. 74073290 residues

Total number of hits satisfying chosen parameters: 522463

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

post-processing: Minimum Match 0%
Maximum Match 100%
listing first 45 summaries

```

Database :
A_Geneseq_ll01.*
1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:*

```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	567	100.0	534	22	AA040125	Human polypeptide
2	567	100.0	534	22	AA092412	Human protein sequ
3	567	100.0	534	22	AA0933478	Human protein sequ
4	567	100.0	534	22	AA0951520	Human protein sequ
5	87.5	15.4	467	21	AA058505	Arabidopsis thalia
6	74	13.1	383	20	AA013942	Human transmembran
7	74	13.1	410	22	AA075511	Human colon cancer
8	70.5	12.4	287	19	AA051088	Coxidia ruminantium
9	70.5	12.4	287	21	AA036182	Coxidia ruminantium
10	70.5	12.4	287	22	AA040192	Major antigenic pr
11	69.5	12.3	1161	21	AA052037	A. aegyptius AAE050

12	69.5	12.3	1161	21	AAV51666	A. aeolicus AASEQ5
13	67	11.8	1256	13	AAW27746	Muramidase release
14	66	11.6	336	14	AAW98449	H. pylori GHP0 578
15	66	11.6	1254	11	AAW07503	Merozoite apical-en
16	66	11.6	1254	18	AAW24575	Merozoite apical-e
17	65.5	11.6	579	21	AAV95702	Cosmid CHRIM5 enco
18	65	11.5	221	21	AAK10203	Arabidopsis thalia
19	65	11.5	221	21	AAK431196	Arabidopsis thalia
20	59	11.5	237	21	AAK10202	Arabidopsis thalia
21	55	11.5	237	21	AAK431195	Arabidopsis thalia
22	55	11.5	278	21	AAV68976	Arabidopsis thalia
23	55	11.5	389	22	AAAB36611	CpsK protein whic
24	64.5	11.4	228	21	AAK49247	Human FLEXH-33 pr
25	64.5	11.4	238	21	AAK18595	Arabidopsis thalia
26	64.5	11.4	238	21	AAK49254	Arabidopsis thalia
27	64.5	11.4	412	20	AAW68010	S. frugiperda immu
28	64.5	11.4	1073	18	AAW32063	Human ST receptor
29	54.5	11.4	1073	13	AAW37371	Human ST receptor
30	53.5	11.2	243	22	AAW82986	S. epidermidis ope
31	63.5	11.2	355	22	AAW83039	S. epidermidis ope
32	62.5	11.0	296	19	AAW98855	H. pylori GHP0 163
33	62.5	11.0	403	20	AAV36952	Chlamydia trachoma
34	62.5	11.0	718	18	AAW01537	Bovine ribonucleas
35	62	10.9	276	18	AAW20832	H. pylori secreted
36	62	10.9	990	14	AAW39343	EpilB protein, Sta
37	52	10.9	990	20	AAV43431	S. epidermis readi
38	62	10.9	1222	21	AAW01828	H. influenzae stra
39	62	10.9	1228	21	AAV96212	Haemophilus infue
40	61	10.7	702	21	AAV956212	P. multocida hyalu
41	60.5	10.7	490	16	AAW25566	Penicillium purpur
42	60.5	10.7	633	15	AAW56696	CryIIA insecticida
43	60.5	10.7	634	21	AAW94259	Cry2A delta-endot
44	60.5	10.7	816	20	AAV37339	Group B Streptoco
45	60.5	10.7	1328	21	AAW349342	Arabidopsis thalia

ALIGNMENTS

RESULT	1
AAH40125	
ID	AAH40125 standard; Protein: 634 AA.
XX	
XX	
AAH40125:	
XX	
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polypeptide SEQ ID NO 3270.
XX	
KW	Human; nootropic; immunosuppressant;
KW	peripheral nervous system; neuropathy;
KW	Alzheimer's; Parkinson's disease; Huntington's; amyotrophic lateral sclerosis; Shy-Drager syndrome; thrombolytic; drug screening; leukaemia.
XX	
XX	
OS	Homo sapiens.
XX	
PN	WO200153312-A1.
XX	
XX	
PD	26-JUL-2001.

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

```

PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yan Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac PI;
XX
DR WPI: 2001-442253/47.
DR N-PSDB: AA159281.
XX
PI Novel nucleic acids and polypeptides, useful for treating disorders
PI such as central nervous system injuries -
XX
PS Example 5: SEQ ID NO 3270; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA151369) and
CC the encoded polypeptides (AA138642-AA142213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 534 AA:

Query Match 100.0%; Score 567; DB 22; Length 534;
Best Local Similarity 100.0%; Pred. No. 5.6e-63;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

QY 1 MSLVPATNYIYIPLNOLKGGTIVNVYGVVKFKPPYLSKGTDCSVVTIVDQTNVKLTCL 50
DB 1 mslvpacnyitpnlqkgtcivnvvgvkvfkppylskgtcysvvtivdqcnvklctcl 50
QY 61 LFSGNYEALPIIYKNGDIVFHRKLTQVYKKETQGITSSGFASLTFEGT 109
DB 61 lfsqnyealpiiykngdivrfrhrlkiqvykktqgitssgfaslftegt 109

RESULT 2
AAB92742
ID AAB92742 standard; Protein: 534 AA.
AC AAB92742:
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:11188.
XX
KW Human: primer: detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074517-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PP 29-JUL-1999; 92JP-0248036.
PP 27-AUG-1999; 99JP-0300253.
PP 11-JAN-2000; 2000JP-0118776.
PP 02-MAY-2000; 2000JP-0183767.
PP 09-JUN-2000; 2000JP-0241859.
XX

```

```

PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PI Primer sets for synthesizing polynucleotides, particularly the 5602
PI full-length cDNAs defined in the specification, and for the detection
PI and/or diagnosis of the abnormality of the proteins encoded by the
PI full-length cDNAs -
XX
PS Claim 9: SEQ ID 11188; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 534 AA:

Query Match 100.0%; Score 567; DB 22; Length 534;
Best Local Similarity 100.0%; Pred. No. 5.6e-63;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

QY 1 MSLVPATNYIYIPLNOLKGGTIVNVYGVVKFKPPYLSKGTDCSVVTIVDQTNVKLTCL 50
DB 1 mslvpacnyitpnlqkgtcivnvvgvkvfkppylskgtcysvvtivdqtnvklctcl 50
QY 61 LFSGNYEALPIIYKNGDIVFHRKLTQVYKKETQGITSSGFASLTFEGT 109
DB 61 lfsqnyealpiiykngdivrfrhrlkiqvykktqgitssgfaslftegt 109

RESULT 3
AAB93478
ID AAB93478 standard; Protein: 534 AA.
XX
AC AAB93478:
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12761.
XX
KW Human: primer: detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074517-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX

```

```

FR 29-JUL-1999: 99JP-0248036.
PR 27-AUG-1999: 99JP-0300253.
PR 11-JAN-2000: 2000JP-0118776.
PR 02-MAY-2000: 2000JP-0183767.
PR 09-JUN-2000: 2000JP-0241899.
XX
PR (HELI-) HELIX RES INST.
PA
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PI full-length cDNAs defined in the specification, and for the detection
PI and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs.
XX
PS Claim 8: SEQ ID 12751: 2537pp + CD ROM: English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primers sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92445 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 534 AA:
SQ
Query Match 100.0%; Score 567; DB 22: Length 634;
Best Local Similarity 100.0%; Pred. No. 5.6e-63;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MSLVPATNYIYTPNLQKGGTIVNVYGVKFFPPYLSKGTDCSVVTVVDQTNVKLTCL 60
DB 1 mslvpatnyiytpnlqkggtivnvgyvkvffkppylskgtidcsvvtvtdqtnvklc! 60
OY 61 LFSGNYEALPIIYKNGDIVRFHRLKIQVYKKEQTGGITSSGFSALTFEGT 109
DB 61 lfsngyealpiiykngdivrfhrkikiqvyykktggtitssgsfalsltfegt 109
RESULT 4
AAB95120
ID AAB95120 standard; Protein: 534 AA.
XX
XX AAB95120;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:17113.
XX
XX Human; primer: detection; diagnosis: antisense therapy; gene therapy.
XX
XX Homo sapiens.
XX
XX

```

```

FN EPI074517-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000: 2000EP-0116126.
XX
XX 22-JUL-1999: 99JP-0248036.
XX 27-AUG-1999: 99JP-0300253.
XX 11-JAN-2000: 2000JP-0118776.
XX 02-MAY-2000: 2000JP-0183767.
XX 09-JUN-2000: 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PI full-length cDNAs defined in the specification, and for the detection
PI and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs.
XX
PS Claim 8: SEQ ID 17113: 2537pp + CD ROM: English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primers sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 534 AA:
SQ
Query Match 100.0%; Score 567; DB 22: Length 634;
Best Local Similarity 100.0%; Pred. No. 5.6e-63;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MSLVPATNYIYTPNLQKGGTIVNVYGVKFFPPYLSKGTDCSVVTVVDQTNVKLTCL 60
DB 1 mslvpatnyiytpnlqkggtivnvgyvkvffkppylskgtidcsvvtvtdqtnvklc! 60
OY 61 LFSGNYEALPIIYKNGDIVRFHRLKIQVYKKEQTGGITSSGFSALTFEGT 109
DB 61 lfsngyealpiiykngdivrfhrkikiqvyykktggtitssgsfalsltfegt 109
RESULT 5
AAG58505
ID AAG58505 standard; Protein: 467 AA.
XX
XX AAG58505;
XX
XX 18-OCT-2000 (first entry)
XX
XX

```

DE Arabidopsis thaliana protein fragment SEQ ID NO: 75533.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX EP1033405-A2.
PN
PD
XX 06-SEP-2000.
XX
PF 25-FEB-2000: 2000EP-0301439.
XX
PR 25-FEB-1999: 99US-0121825.
PR 05-MAR-1999: 99US-0123180.
PR 09-MAR-1999: 99US-0123548.
PR 23-MAR-1999: 99US-0125788.
PR 25-MAR-1999: 99US-0125254.
PR 29-MAR-1999: 99US-0125785.
PR 01-APR-1999: 99US-0127462.
PR 05-APR-1999: 99US-0128234.
PR 08-APR-1999: 99US-0128714.
PR 16-APR-1999: 99US-0129845.
PR 19-APR-1999: 99US-0130077.
PR 21-APR-1999: 99US-0130449.
PR 23-APR-1999: 99US-0130510.
PR 28-APR-1999: 99US-0130891.
PR 30-APR-1999: 99US-0131449.
PR 30-APR-1999: 99US-0132048.
PR 30-APR-1999: 99US-0132407.
PR 04-MAY-1999: 99US-0132484.
PR 05-MAY-1999: 99US-0132485.
PR 06-MAY-1999: 99US-0132486.
PR 07-MAY-1999: 99US-0132487.
PR 11-MAY-1999: 99US-0132863.
PR 14-MAY-1999: 99US-0134256.
PR 14-MAY-1999: 99US-0134318.
PR 14-MAY-1999: 99US-0134219.
PR 14-MAY-1999: 99US-0134221.
PR 14-MAY-1999: 99US-0134370.
PR 18-MAY-1999: 99US-0134768.
PR 19-MAY-1999: 99US-0134941.
PR 20-MAY-1999: 99US-0135124.
PR 21-MAY-1999: 99US-0135353.
PR 24-MAY-1999: 99US-0135529.
PR 25-MAY-1999: 99US-0136021.
PR 27-MAY-1999: 99US-0136392.
PR 28-MAY-1999: 99US-0136782.
PR 01-JUN-1999: 99US-0137222.
PR 03-JUN-1999: 99US-0137528.
PR 04-JUN-1999: 99US-0137502.
PR 07-JUN-1999: 99US-0137724.
PR 08-JUN-1999: 99US-0138094.
PR 10-JUN-1999: 99US-0138540.
PR 10-JUN-1999: 99US-0138847.
PR 14-JUN-1999: 99US-0139119.
PR 16-JUN-1999: 99US-0139452.
PR 16-JUN-1999: 99US-0139453.
PR 17-JUN-1999: 99US-0139492.
PR 18-JUN-1999: 99US-0139454.
PR 18-JUN-1999: 99US-0139455.
PR 18-JUN-1999: 99US-0139456.
PR 18-JUN-1999: 99US-0139457.
PR 18-JUN-1999: 99US-0139458.
PR 18-JUN-1999: 99US-0139459.
PR 18-JUN-1999: 99US-0139460.
PR 18-JUN-1999: 99US-0139461.
PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139753.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139899.
PR 23-JUN-1999: 99US-0140351.
PR 24-JUN-1999: 99US-0140595.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 02-JUL-1999: 99US-0142154.
PR 06-JUL-1999: 99US-0142055.
PR 08-JUL-1999: 99US-0142390.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142977.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 19-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144352.
PR 20-JUL-1999: 99US-0144632.
PR 20-JUL-1999: 99US-0144884.
PR 21-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145088.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145089.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.
PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145918.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 02-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 25-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151060.

Fri Apr 12 08:54:22 2002

```

XX PD 05-APR-2001.
XX PF
XX PR 28-SEP-2000: 2000WO-US26524.
XX PR 29-SEP-1999: 99US-0157137.
XX PR 03-NOV-1999: 99US-0163280.
XX PR (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA:
XX PI WPI: 2001-235357/24.
XX DR N-PSDB; AAH35016.
XX XX
XX XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX PT
XX PS Claim 11: Page 7845-7847: 9803pp: English.
XX XX
XX XX AAH32943 to AAH37195 and AAH37514 to AAH37788 represent human colon
XX CC cancer-associated nucleic acid molecules (N) and proteins (P), where
XX CC the proteins are collectively known as colon cancer antigens. The colon
XX CC cancer antigens have cytostatic activity and can be used in gene
XX CC therapy and vaccine production. N and P may be used in the prevention,
XX CC diagnosis and treatment of diseases associated with inappropriate P
XX CC expression. For example, N and P may be used to treat disorders
XX CC associated with decreased expression by rectifying mutations or deletions
XX CC in a patient's genome that affect the activity of P by expressing
XX CC inactive proteins or to supplement the patient's own production of P.
XX CC Additionally, N may be used to produce the colon cancer-associated Ps
XX CC by inserting the nucleic acids into a host cell and culturing the cell
XX CC to express the proteins. N and P can be used in the prevention, diagnosis
XX CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX CC and AAH37789 represent sequences used in the exemplification of the
XX CC present invention.
XX CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX CC missing at time of publication, meaning no sequences are present for
XX CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX XX
XX S0 Sequence 410 AA:

Query Match 13.1%; Score 74; DB 22; Length 410;
Best Local Similarity 26.2%; Pred. No. 0.82; Mismatches 4; Gaps 2;
Matches 17; Conservative 18; Indels 26;

Oy 36 YLSKGDYCSVVIIVDQIRYKL--TCLLFSNGYEALPIIYK--NGDIVRPHRLKIOVYKK 91
Db 275 hlsfgedypgivnpldhtn:tapqasmfgyfkvvpvymkvdge:lrtngqfsvt:hek 334

Oy 92 ETOGI 96
Db 335 vangl 339

RESULT 8
AAW51088 standard: Protein; 287 AA.
XX AC AAW51088;
XX XX
XX DT 14-SEP-1998 (first entry)
XX DE Cowdria ruminantium major antigenic protein 1 (MAP1).
XX XX
XX KW MAP1 gene; major antigenic protein 1; rickettsia; heartworm;
XX KW DNA vaccine.
XX XX
XX CS Cowdria ruminantium.
XX XX
XX PN WC2816554-A1.
XX XX

```

```

PD 23-APR-1998.
XX 17-OCT-1997: 97WO-US19044.
XX PF
XX PR 17-OCT-1996: 96US-0733230.
XX PA (UYFL) UNIV FLORIDA.
XX PI Barbet AF, Burridge MJ, Canta RR, Mahan SM, McGuire TC:
XX PI Nyika A, Rurangirwa FR:
XX PI WPI: 1998-251232/22.
XX DR N-PSDB; AAV07176.
XX XX
XX XX Composition containing nucleic acid encoding rickettsial antigen -
XX PT useful for, e.g. stimulating protective immune response in humans or
XX PT animals
XX XX
XX PS Claim 3: Page 15-16: 39pp: English.
XX XX
XX XX This polypeptide comprises the major antigen protein 1 gene (MAP1)
XX CC of Cowdria ruminantium, the causative agent of heartwater in
XX CC domestic ruminants. It is encoded by the MAP1 gene (see AAV07176).
XX CC A claimed composition comprises a nucleic acid (see AAV07176-82)
XX CC encoding a polypeptide (see AAW51088-99) that elicits a protective
XX CC immune response against a rickettsial pathogen. The nucleic acid
XX CC is used, in human or veterinary medicine, in vaccines to protect
XX CC against Rickettsia, Ehrlichia, Anaplasma and Cowdria species. The
XX CC nucleic acid does not replicate in the host but remains episomal
XX CC and capable of expressing polypeptide for at least 19 mth.
XX XX
XX S0 Sequence 287 AA:

Query Match 12.4%; Score 70.5; DB 19; Length 287;
Best Local Similarity 30.3%; Pred. No. 1.4;
Matches 27; Conservative 19; Mismatches 18; Indels 25; Gaps 7;

Oy 35 PYLSK--GTDYCSVVIIVDQIRYKLTC-----LLFSNGYEALPIIYKNGDIVRPHRL--- 84
Db 194 pyvcagiqtd---lvsvinatnplksyqgklgisysinsea--sifigg---hfhrrvign 245

Oy 85 -----KIOVYKKETOGITSSGFASLIFE 107
Db 246 efkdlatikfiskskt-gisnppafasatid 273

RESULT 9
AAB36182 standard: Protein; 287 AA.
XX AC AAB36182;
XX XX
XX DT 02-MAR-2001 (first entry)
XX DE Cowdria ruminantium MAP1.
XX XX
XX KW Cowdria ruminantium; MAP1; major antigenic protein 1; antirickettsial;
XX KW vaccine; gene therapy; Rickettsia; Ehrlichia; Anaplasma; map2; lhworf3;
XX KW 4hworfl; 18hworfl; 3gdorfl.
XX XX
XX OS Cowdria ruminantium.
XX XX
XX PN WC200005063-A2.
XX XX
XX PD 02-NOV-2000.
XX XX
XX PE 21-APR-2000: 2000WO-US10886.
XX PF
XX XX 22-APR-1999: 99US-0130725.
XX XX
XX XX (UYFL) UNIV FLORIDA.
XX XX

```

PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FP, Moreland AL, Simbi BH, Whitmire KW, Alleman AP;
 XX WPI: 2000-579575/66.
 DR N-PSDB: AAC58699.
 XX
 XX New polynucleotides useful as DNA vaccines for conferring immunity to
 PI rickettsial infection e.g. heartwater caused by Cowdria ruminantium.
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
 PT -
 XX
 PS Claim 3: Page 33: 53pp: English.
 XX
 XX The present sequence is given in a specification relating to nucleic
 CC acid vaccines containing genes to protect animals or humans against
 CC rickettsial diseases caused by a organisms of Rickettsia sp., Ehrlichia
 CC sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The vaccine comprises the
 CC major antigenic protein 1 (MAP1) gene or the major antigenic protein 2
 CC (MAP2) gene of rickettsial pathogens. The nucleic acid vaccines can be
 CC driven by the human cytomegalovirus (HCMV) enhancer-promoter. Cowdria
 CC ruminantium genes designated map 2, lhwoif3, 4hwoif1, 18hwoif1 and
 CC 3gdoif3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.
 XX
 SO Sequence 287 AA:
 Query Match 12.4% Score 70.5; DB 21: Length 287;
 Best Local Similarity 30.3%; Pred. No. 1.4;
 Matches 27: Conservative 19: Mismatches 18: Indels 25: Gaps 7:
 Oy 35 PYLSK--GIDYCSVVTIVDQTNVRLTC-----LLFSGNYEALPIIYKNGDIVRFHRL--- 84
 ||| ||| :|||: || ||| : || ||| ||| |||
 Db 194 pycagigtgtd---lvsinatnplksyqgklgisysinsea--sifigg---hfhrrvign 215
 Oy 85 -----KIOVYKKEITOGITSSGFASLIFE 107
 ||| :|||: ||| ||| : || ||| ||| |||
 Db 246 efkdia tikiftskt-gisnpgfasatid 273
 RESULT 10
 AAU04192
 ID AAU04192 standard; Protein: 287 AA.
 XX
 AC AAU04192:
 XX
 DT 23-OCT-2001 (first entry)
 XX
 DE Major antigenic protein 1 (MAP1).
 XX
 KW Major antigenic protein 1: MAP1; vaccine: immunogenic; rickettsia;
 KW infection; heartwater; diagnostic.
 XX
 OS Cowdria ruminantium.
 XX
 PN US6251872-B1.
 XX
 PD 26-JUN-2001.
 XX
 PF 17-OCT-1997: 97US-0953326.
 XX
 PR 17-OCT-1996: 96US-0733230.
 XX
 XX (UYEL) UNIV FLORIDA.
 PA Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
 PI Rurangirwa FP, Mahan SM, Bowie MV, Alleman AP;
 XX WPI: 2001-424487/45.
 DR N-PSDB: AAS07575.
 CC

XX New MAP2 genes and polypeptides useful as vaccines for conferring
 PI immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PI as molecular markers in nucleic acid analysis procedures
 XX
 XX Example 1: Column 11-13: 30pp: English.
 XX
 XX The sequence represents the amino acid sequence of major antigenic
 CC protein 1 (MAP1) isolated from Cowdria ruminantium. The MAP
 CC polynucleotides and polypeptides are useful as vaccines for conferring
 CC immunity to rickettsia infection, including Cowdria ruminantium causing
 CC heartwater. The MAP polynucleotides may be used as molecular markers in
 CC nucleic acid analysis procedures, and to produce the MAP polypeptides,
 CC which may be used to raise antibodies that are reactive with the
 CC polypeptides. The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious
 CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is
 CC needed.
 XX
 SO Sequence 287 AA:
 Query Match 12.4% Score 70.5; DB 22: Length 287;
 Best Local Similarity 30.3%; Pred. No. 1.4;
 Matches 27: Conservative 19: Mismatches 18: Indels 25: Gaps 7:
 Oy 25 PYLSK--GIDYCSVVTIVDQTNVRLTC-----LLFSGNYEALPIIYKNGDIVRFHRL--- 84
 ||| ||| :|||: || ||| : || ||| ||| |||
 Db 194 pycagigtgtd---lvsinatnplksyqgklgisysinsea--sifigg---hfhrrvign 245
 Oy 85 -----KIOVYKKEITOGITSSGFASLIFE 107
 ||| :|||: ||| ||| : || ||| ||| |||
 Db 246 efkdia tikiftskt-gisnpgfasatid 273
 RESULT 11
 AAU52037
 ID AAU52037 standard; Protein: 1161 AA.
 XX
 AC AAU52037:
 XX
 DT 12-JUL-2000 (first entry)
 XX
 DE A. aeolicus AASEQ50 protein.
 XX
 KW Polymerase; thermostable; nucleic acid elongation; sliding clamp protein;
 KW amplification; reverse transcription.
 XX
 CS Aquifex aeolicus.
 XX
 PN DE19840771-A1.
 XX
 PD 10-FEB-2000.
 XX
 PF 07-SEP-1998: 98DE-1040771.
 XX
 PP 06-AUG-1998: 98DE-1035653.
 XX
 FA (LICR-) LION BI-SCIENCE AG.
 XX
 PI Voss H, Moekkel G, Kober J, Kildner C;
 DR WPI: 2000-207142/19.
 XX
 XX A thermostable in vitro polymerase complex for template-dependent
 PI elongation of nucleic acids in amplification or reverse transcription
 PI methods
 XX
 PS Disclosure: Page 129-134; 152pp: German.
 XX
 CC This invention describes a novel thermostable in vitro complex for

CC template-dependent elongation of nucleic acids which comprises a
 CC thermostable sliding clamp protein, which is connected with a
 CC elongation protein that shows thermostable polymerase activity. The
 CC thermostable in vitro accessory complex can be used to produce the
 CC thermostable in vitro complex, which is useful for template-dependent
 CC elongation of nucleic acids, e.g. for amplification or reverse
 CC transcription. This is useful for sequencing nucleic acids by the
 CC polymerase chain reaction or reverse transcriptase-polymerase chain
 CC reaction (RT-PCR). The complex can be used to mark nucleic acids.
 CC AAY52000-Y52084 and AAY90752-W90799 represent proteins and protein
 CC fragments used to illustrate the method of the invention.
 XX
 SQ Sequence 1161 AA:
 Query Match 12.3% Score 69.5; DB 21: Length 1161;
 Best Local Similarity 29.3%; Pred. No. 13;
 Matches 22: Conservative 12: Mismatches 24: Indels 17: Gaps 4:
 Oy 5 PATNY-----IYELNOLK-----GGTIVNVYGVVKKFPYLSKGTIVSVVITVDQ 52
 Db 972 pldnyekllknytpiedleekeseavltgvtelk:kk-----tkngdy-mavfnlvdK 1027
 Oy 53 INVKLTCLLFSGNYE 67
 Db 1028 tgl-icrrvfpgye 1041
 Query Match 12.3% Score 69.5; DB 21: Length 1161;
 Best Local Similarity 29.3%; Pred. No. 13;
 Matches 22: Conservative 12: Mismatches 24: Indels 17: Gaps 4:
 Oy 5 PATNY-----IYELNOLK-----GGTIVNVYGVVKKFPYLSKGTIVSVVITVDQ 52
 Db 972 pldnyekllknytpiedleekeseavltgvtelk:kk-----tkngdy-mavfnlvdK 1027
 Oy 53 INVKLTCLLFSGNYE 67
 Db 1028 tgl-icrrvfpgye 1041
 RESULT 12
 ID AAY51566
 XX AAY51566 standard; protein; 1161 AA.
 AC AAY51566:
 XX
 DI 01-JUN-2000 (first entry)
 DE A. aeolicus AASE050 protein fragment.
 KW Thermotable; template-dependent elongation; staple protein;
 KW elongation protein; amplification; reverse transcription.
 XX Aquifex aeolicus.
 OS
 PN W0200008164-A2.
 XX
 CD 17-FEB-2000.
 XX
 PF 06-AUG-1999: 99WO-DF02480.
 XX
 PP 06-AUG-1998: 99DE-1035653.
 PP 07-SEP-1998: 98DF-1040771.
 PP 18-JUN-1993: 99EP-0111795.
 XX
 PA (LION-) LION BIOSCIENCE AG.
 XX
 PI Kilger C, Kober I, Voss H, Moeckel G;
 XX
 DE WPI: 2000-195576/17.
 XX
 PI Thermotable in vitro polymerase complex for template-dependent
 PI elongation of nucleic acids
 XX
 PS Disclosure: Page 223-227: 233pp; German.
 XX
 CC This invention describes a novel thermostable in vitro complex (:) for
 CC template-dependent elongation of nucleic acids (NA) comprising a
 CC thermostable "staple" protein and an elongation protein. The thermostable
 CC in vitro complex is useful for template-dependent elongation of NA, e.g.
 CC for amplification or reverse transcription. This is useful for sequencing
 CC NA by the polymerase chain reaction or reverse transcriptase-PCR. The
 CC complex can be used to mark NA. This sequence represents the Aquifex
 CC aeolicus AASE050 protein which is used to illustrate the method of the
 CC invention.

XX SQ Sequence 1161 AA:
 Query Match 12.3% Score 69.5; DB 21: Length 1161;
 Best Local Similarity 29.3%; Pred. No. 13;
 Matches 22: Conservative 12: Mismatches 24: Indels 17: Gaps 4:
 Oy 5 PATNY-----IYELNOLK-----GGTIVNVYGVVKKFPYLSKGTIVSVVITVDQ 52
 Db 972 pldnyekllknytpiedleekeseavltgvtelk:kk-----tkngdy-mavfnlvdK 1027
 Oy 53 INVKLTCLLFSGNYE 67
 Db 1028 tgl-icrrvfpgye 1041
 RFSULT 13
 ID AAR27746
 XX AAR27746 standard; Protein: 1256 AA.
 AC AAR27746:
 XX
 DI 04-MAR-1993 (first entry)
 DE Muramidase released protein.
 KW MRP: detection; prevention; screening; diagnostic.
 XX Streptococcus suis type II pathogenic.
 OS
 FH Key Location/Qualifiers
 FI Peptide 1...47
 FI Peptide /note= "signal peptide"
 FI Peptide 48...1256
 FI Peptide /note= "mature peptide"
 FI Region 857...952
 FI Peptide /note= "proline-rich region"
 FI Peptide 1222...1256
 FI Peptide /note= "membrane anchor sequence"
 XX WC2215630-A.
 XX
 PD 01-OCT-1992.
 XX
 PF 19-MAR-1992: 92WU-RL00054.
 XX
 PP 21-MAR-1991: 91HL-0000510.
 XX
 PR (DIFP-) CENT DIERGENESKUNDIG INST.
 XX
 PA Smith HE, Vecht U;
 XX
 DR WPI: 1992-349215/42.
 DR R-PSDB: AAO29472.
 XX
 PI Deoxyribonucleic acid encoding virulence characteristic of
 PI Streptococcus suis - useful for antibody and polypeptide for
 PI diagnosing and preventing infections in pigs and humans
 XX
 FS Claim 9: Fig 2: 85pp; English.
 XX
 CC The sequence is that of the muramidase released protein from
 CC Streptococcus suis type II (pathogenic) which allows the detection
 CC and the prevention of infections by S. suis in a more effective
 CC manner than was previously possible. It facilitates screening of
 CC e.g. pigs and elimination of infected and carrier pigs can then be
 CC carried out. The new diagnostic tests can distinguish between
 CC a virulent and avirulent strains. It may be used in the prodn. of
 CC a vaccine. See also AAR27744 and AAR27745.
 XX
 SQ Sequence 1256 AA:

```

Query Match      11.8%; Score 67; DB 13; Length 1256;
Best Local Similarity 23.2%; Pred. No. 29;
Matches 28; Conservative 15; Mismatches 13; Indels 20; Gaps 5;

OY 7 TNYIYPLNOLKGGTIVHVVY-----GVVKFKP----PYLSKGTGYCSVVTIVDGIHWKL 57
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1000 tthlykktpevkngtvyvnyvtedgtvik--epvtdtptspgtpy-----dt'unkp 1050

OY 58 TCLLFSG-NYEALPII-YKNGDIVRFRHLKIQVYKK 91
   :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1051 kltlfkgeyelvrldgtengkvvegtvvtvyvrk 1086

RESULT 14
AAW98449
ID AAW98449 standard; Protein: 335 AA.
XX
AC
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 578 protein.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
XX
OS Helicobacter pylori.
XX
FN WO9843478-Al.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCJ INC.
PA (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR WPI: 1998-542293/46.
DR N-PSDB: AAX14169.
XX
PI New isolated Helicobacter polynucleotides - used to develop products
PI for the diagnosis, prevention and treatment of Helicobacter
PI infections and gastrointestinal diseases
XX
PS Claim 8; Page 781-783; 2054pp; English.
XX
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 335 AA:

```

```

Query Match      11.6%; Score 66; DB 19; Length 336;
Best Local Similarity 35.4%; Pred. No. 55;
Matches 17; Conservative 8; Mismatches 13; Indels 10; Gaps 2;

OY 31 FFKPPYLSKGTGYCSVVTIVDQINVKLTCLLFSGNVEALPIIYKNGDI 78
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 58 ffgppglgk-tsishiinkemetnikit-----aapmicksaal 95

```

```

RESULT 15
AAW07503
ID AAR07503 standard; protein: 1254 AA.
XX
AC AAR07503;
XX
DI 06-FEB-1991 (first entry).
XX
DE Merozoite apical-end-localised protein (MAEP) insert 5.3.
XX
KW Malaria; vaccine; Duffy blood group.
XX
OS Plasmodium vivax.
XX
FN WO9011772-A.
XX
PI 18-OCT-1990.
XX
PF 03-APR-1990; 90WO-US01849.
XX
PR 06-APR-1989; 89US-0334270.
PR 05-APR-1989; 89US-0334041.
XX
PA (UYNY-) NEW YORK UNIV.
XX
PI Barnwell JW, Galinski MR, Wertheimer SP;
XX
WPI: 1990-334616/44.
DR N-PSDB: AAQ06328.
XX
PI Malarial apical end merozoite proteins and peptide(s) - used for
PI developing cpds. for treating, preventing and diagnosing malarial
PI infection
XX
PS Claim 2; Fig 1a; 56pp; English.
XX
CC A MAEP compound having a binding affinity for a Duffy blood group
CC antigen of primate red blood cells, is antigenic for the complete
CC protein, and may be used in diagnosis, treatment and vaccination
CC against invasion by P.vivax and P.knowlesi.
XX
SQ Sequence 1254 AA:

Query Match      11.6%; Score 66; DB 11; Length 1254;
Best Local Similarity 26.3%; Pred. No. 39;
Matches 25; Conservative 11; Mismatches 28; Indels 34; Gaps 4;

OY 8 NYIYELPLNOLKG--GTIVNVYGVVKFKPPYLSKGTGYCSVVTIVDQINVKLTCLLFSGN 65
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 491 nlivteknrlngldstltiniegalkesk-----gn 520

OY 55 YEALPIIYKNGDIVRFRHLKIQVYKKETOGITSSGFASL 104
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DI 521 ye-igtlekleelgknrkikvdkitkslms-tvgnfssl 557

Search completed: April 12, 2002. 08:35:17
Job time: 121 sec

```

|||||
541 vfmftfildgtvleaylmsdskffqipaevlmdldlqkgydmldmfcppgikiday 600
601 pwlecfikswnvngtdnqicyciofddtvaedvi 634
601 pwlecfikswnvngtdnqicyciofddtvaedvi 634

RESULT 2
AAM40125
ID AAM40125 standard; Protein: 634 AA.
AC AAM40125:
DT 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 3270.
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia.
OS Homo sapiens.
PN WO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0691036.
XX 29-NOV-2000; 2000US-0727344.
(HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Mu Y, Qian XB, Ren F, Wang D;
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
XX N-PSDB; AAI59281.
XX Novel nucleic acids and polypeptides, useful for treating disorders
such as central nervous system injuries.
XX Example 5: SEQ ID NO 3270; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
the encoded polypeptides (AAM38642-AAM42213) with nootropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: Immune system suppression,
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening,
assays for receptor activity, arthritis and inflammation, leukaemias and
C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
specification.

full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -
C1a1m 8; SEQ ID 11188; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesising 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesising polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
cDNAs easily without any specialised methods. AAI03166 to AAI13628 and
AAI13633 to AAI18742 represent human cDNA sequences; AAB92446 to
AAB95893 represent human amino acid sequences; and AAI13629 to AAI13632
represent oligonucleotides, all of which are used in the exemplification
of the present invention.

Query Match 100.0%; Score 3324; DB 22; Length 634;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 634; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSLVPATNYITPLNQLKGGTIVNMGVWFFKPPYLSKGTDCYSVTVVQDNVKTCL 60
DB 1 mslvpatnyitplnqlkgtivnmgvwwffkppylskgtdcysvvtivdqnvktcl 60
QY 61 LPSGNYEALPIYKNGDIVRPHRIKIOYKKETQGTSSGFSALTFEGTLAGAPLPTSS 120
DB 61 lpsgnyealpiyknqdivrphrikioykykktqgtssgfsaltfegtlagaplptss 120
QY 121 KYENFTTDDHKMVAALRWASHTMSPSWTLLKLDVQPMQYFDTICOLLAKAEVDCASPL 180
DB 121 kyentfttdhkmvwalrwasthmspswtllkldvqpmqyfditcollakaevdgasfl 180
QY 181 LKVMGTRTPPSWRVLQDLVLEGLSHIHRLQNLITDILVDNIIVIVARSLSKVGSLR 240
DB 181 lkvmgtrtppswrvlqdlvleghshihrlqnlitdilvynihvhsarslkvgslr 240
QY 241 IYSLHTKLOSMMSENQTMISLFFHLHGGTSYGRGIRVLPSNSVDOLKKDLKESANI.TAN 300
DB 241 iyslhtklosmmseentmlslffhlhggtsygrgirvlpsnsvdolkldlesanitan 300
QY 301 QHSDVLCQSEPDSPSSSGSVSYEVERCQOQLSATLTIDHQYLERTPLCALIKKAKVQOY 360
DB 301 qhsdvlcqsepdspssgsvsyevercqqlsatltldhgylerptlcalikakvqptqy 360
QY 361 RIRAKLSRYKPRRLFOVKLHCPKCHLLOEVPHEGDLIDIFODGATKTPPVKLQNTSLYD 420
DB 361 riraklsrykprrlfsvklhpcpkchllqevphegdlidifodgatktpvkvkntslslyd 420
QY 421 SKTWTNKGKRVAVHFKVKNNGILPLNSNECLLLTEGTLSEICKLSKNKFSNIPVRSCH 480
DB 421 sktwtknkgkrvavhfkvknngilplnsnecllltegtlseicklsknkfsnvpvrsgh 480
QY 481 EDLELIDLSAPFLIOQTIIHYGCKQCSLSRSIQNLNSLVKDTSWITPSSVAALGIVPQY 540
DB 481 edlelidlsapfliqgtiihygckqcsllrsiqnlinslvkdtswitpssvaealgivpqy 540
QY 541 VFVWTFLLDGDGTGLAYLMSDSKFFQIPASEVLMDDDLQKSDVMDIMDMFCPPGKIDAY 600

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: Apr 11 9, 2002, 16:58:21 ; Search time 58.75 seconds
(without alignments) 799.361 Million cell updates/sec

Title: US-09-816-248-13
Sequence: 1 MSUVATATVITVPLNOLKGC.....GTDNOCVQIFPTVAHEVY 634

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SID8/gcgdata/geneseq/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3324	100.0	634	22	AAB92742
2	3317	99.8	634	22	AAAM40125
3	3317	99.8	634	22	AAAB34478
4	3317	99.8	634	22	AAAB95120
5	1589	47.8	307	22	AAAM1911
6	358	10.8	66	22	AAAC73583
7	147	4.4	30	22	AAAM34264
8	138.5	4.2	467	21	AAAG58505
9	122	3.7	1144	17	AAAB8122
10	122	3.7	1144	17	AAAB8123
11	114	3.4	2391	15	AAAB5694

12	111	3.3	1139	21	AAAB18793	Amino acid sequenc
13	106	3.2	1123	21	AAAG35865	Arabidopsis thalia
14	105.5	3.2	1557	21	AAAB84730	Amino acid sequenc
15	105.5	3.2	1574	21	AAAB84729	Amino acid sequenc
16	105.5	3.2	1578	21	AAAB84731	Amino acid sequenc
17	105.5	3.2	1588	21	AAAB84733	Amino acid sequenc
18	105.5	3.2	1605	21	AAAB84733	Amino acid sequenc
19	105	3.2	647	21	AAAB84733	Amino acid sequenc
20	105	3.2	1112	21	AAAB84733	Amino acid sequenc
21	105	3.2	1112	21	AAAB84733	Amino acid sequenc
22	104	3.1	785	21	AAAB84733	Amino acid sequenc
23	103.5	3.1	1002	21	AAAB84733	Amino acid sequenc
24	103.5	3.1	1266	18	AAAB84733	Amino acid sequenc
25	103.5	3.1	1266	18	AAAB84733	Amino acid sequenc
26	103.5	3.1	1568	20	AAAB84733	Amino acid sequenc
27	103.5	3.1	1568	21	AAAB84733	Amino acid sequenc
28	103.5	3.1	1568	21	AAAB84733	Amino acid sequenc
29	103.5	3.1	1568	22	AAAB84733	Amino acid sequenc
30	103.5	3.1	1568	22	AAAB84733	Amino acid sequenc
31	103.5	3.1	1568	22	AAAB84733	Amino acid sequenc
32	102.5	3.1	1311	22	AAAB84733	Amino acid sequenc
33	102.5	3.1	4393	22	AAAB84733	Amino acid sequenc
34	101	3.0	611	22	AAAB84733	Amino acid sequenc
35	101	3.0	1604	19	AAAB84733	Amino acid sequenc
36	100.5	3.0	653	22	AAAB84733	Amino acid sequenc
37	100.5	3.0	1119	20	AAAB84733	Amino acid sequenc
38	100.5	3.0	1119	21	AAAB84733	Amino acid sequenc
39	100.5	3.0	1119	22	AAAB84733	Amino acid sequenc
40	100.5	3.0	1119	22	AAAB84733	Amino acid sequenc
41	100.5	3.0	1119	22	AAAB84733	Amino acid sequenc
42	100.5	3.0	1119	22	AAAB84733	Amino acid sequenc
43	100.5	3.0	1119	20	AAAB84733	Amino acid sequenc
44	100	3.0	1040	21	AAAB84733	Amino acid sequenc
45	100	3.0	1059	21	AAAB84733	Amino acid sequenc

ALIGNMENTS

RESULT 1
AAB92742 standard; Protein: 634 AA.
XX
AC AAB92742:
XX
XX 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:11188.
XX
XX Human: primer: detection; diagnosis; antisense therapy; gene therapy.
OS Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ciba T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ushii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602

```
Db 1 mslypatnnytytplnqkqgtivnvgvkvifkppylskgdcycsvvclvdqcnvklctcl 60
Qy 61 LFSGNYEALPIIYKNGDIVRPHRLKIOVYKKEQTGITSQGFASLTFTGLGAPTIPTSS 120
Db 61 lfsngyealpIiYkngdivrPhrlKioVykKetgItSQGFASLTftGLgApTIptSS 120
Qy 121 KYFNFTTEDIHKMVEALRWASTHMSPSWTLKLCQVPMQYFDLTCLQLGKAEVDCASFL 180
Db 121 kyfnfttedihkmvealrwasthmspswtlklcdvqpmqyfdltcqlgkaevdgsfl 180
Qy 181 LKVDGTRTPPPSRVRLIQDLVLEGDLSHIRLQNLITDILVVDNHVHVARSLKVGSEFLR 240
Db 181 lkvdgtrtpfppsrvrliqdlvlegdshlrlqnlitdilvvdnhvhvarslkvgsflr 240
Qy 241 IYSLHTKLSQMSNENQTMLSLEFHLHGCTSYGRGIRVLPESNSDVQDLKDLLESANLTAN 300
Db 241 IYSLHTKLSQMSNENQTMLSLEFHLHGCTSYGRGIRVLPESNSDVQDLKDLLESANLTAN 300
Qy 301 QHSDVICQSEPDSPSSGVSLSYEVERCOQLSATILTIDHQLERTPLCALILKQKAPQOY 360
Db 301 qhsdvicqsepdspssgvslyevercqqlsatiltidhqlertplcalilkqkapy 360
Qy 361 RIRAKLSYKPRRLFQSVKLHCPKCHLLQEVPHGDLDIIFQDGATKTPDVKLQNTSLYD 420
Db 361 riraklsykprrlfqsvklhcpkchllqevphgdlidifqdgatktpdvklqntsl 420
Qy 421 SKIWTTKNQGRKVAHVHVKNNGLPLSNCELLLEGGLTSEICKLSNKFNSVIPVRSCH 480
Db 421 skiwtcknqgrkvahvhknnlplsncellleggltselcklsnkfnsvipvrsg 480
Qy 481 EDLELDSLAPFLIOGTIIHHYGTGYCT 507
Db 481 edleldlsapfligtthhygckcs 507
```

RESULT 3

AAB95120
ID AAB95120 standard; Protein: 634 AA.

AC AAB95120;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:17113.

XX human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 29-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs.

RESULT 4

AAB92742

ID AAB92742 standard; Protein: 634 AA.

XX

PS Claim 8: SEQ ID 17113; 2537pp + CU' KUM; English.

XX

CC The present invention describes primer sets for synthesizing 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dr primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesizing polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
represent oligonucleotides, all of which are used in the exemplification
of the present invention.

XX Sequence 634 AA;

Query Match 96.5%; Score 2637; DB 22; Length 634;
Best Local Similarity 99.2%; Pred. No. 9.5e-263;
Matches 503; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MSlyPATNnyTYtPlnQkqGtIvNvgVkvIFkppYlSkGdcYcSVvclVdqCnvKlctcl 60
Db 1 mslypatnnytytplnqkqgtivnvgvkvifkppylskgdcycsvvclvdqcnvklctcl 60

Qy 61 LFSGNYEALPIIYKNGDIVRPHRLKIOVYKKEQTGITSQGFASLTFTGLGAPTIPTSS 120
Db 61 lfsngyealpIiYkngdivrPhrlKioVykKetgItSQGFASLTftGLgApTIptSS 120

Qy 121 KYFNFTTEDIHKMVEALRWASTHMSPSWTLKLCQVPMQYFDLTCLQLGKAEVDCASFL 180
Db 121 kyfnfttedihkmvealrwasthmspswtlklcdvqpmqyfdltcqlgkaevdgsfl 180

Qy 181 LKVDGTRTPPPSRVRLIQDLVLEGDLSHIRLQNLITDILVVDNHVHVARSLKVGSEFLR 240
Db 181 lkvdgtrtpfppsrvrliqdlvlegdshlrlqnlitdilvvdnhvhvarslkvgsflr 240

Qy 241 IYSLHTKLSQMSNENQTMLSLEFHLHGCTSYGRGIRVLPESNSDVQDLKDLLESANLTAN 300
Db 241 IYSLHTKLSQMSNENQTMLSLEFHLHGCTSYGRGIRVLPESNSDVQDLKDLLESANLTAN 300

Qy 301 QHSDVICQSEPDSPSSGVSLSYEVERCOQLSATILTIDHQLERTPLCALILKQKAPQOY 360
Db 301 qhsdvicqsepdspssgvslyevercqqlsatiltidhqlertplcalilkqkapy 360

Qy 361 RIRAKLSYKPRRLFQSVKLHCPKCHLLQEVPHGDLDIIFQDGATKTPDVKLQNTSLYD 420
Db 361 riraklsykprrlfqsvklhcpkchllqevphgdlidifqdgatktpdvklqntsl 420

Qy 421 SKIWTTKNQGRKVAHVHVKNNGLPLSNCELLLEGGLTSEICKLSNKFNSVIPVRSCH 480
Db 421 skiwtcknqgrkvahvhknnlplsncellleggltselcklsnkfnsvipvrsg 480

Qy 481 EDLELDSLAPFLIOGTIIHHYGTGYCT 507
Db 481 edleldlsapfligtthhygckcs 507

XX AC AAB92742;
 XX AC 26-JUN-2001 (first entry)
 XX DE Human protein sequence SEQ ID NO:11188.
 XX DE Human; primer: detection; diagnosis; antisense therapy; gene therapy.
 XX OS Homo sapiens.
 XX PN EP1074617-A2.
 XX PD 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-0116126.
 XX PR 29-JUL-1999; 99JP-0248036.
 XX PR 27-AUG-1999; 99JP-0300253.
 XX PR 11-JAN-2000; 2000JP-0118776.
 XX PR 02-MAY-2000; 2000JP-0183767.
 XX PR 09-JUN-2000; 2000JP-0241899.
 XX PA (HELI-) HELIX RES INST.
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nishii K, Otsuki Y;
 XX DR WPI: 2001-318749/34.
 XX PT primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs.
 XX PS Claim 8: SEQ ID 11188; 2537bp + CD ROM; English.
 XX XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human CDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX SO Sequence 634 AA:

Query Match 96.2%; Score 2628; DB 22: Length 634;
 Best Local Similarity 99.0%; Pred. No. 8.1e-262;
 Matches 502; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSVLPATNYTPTPLNGKGTIVNVGVVFPFPPYLSKGTQCSVVTVDQTNVLTCL 60
 DB 1 mslypatnytytptplngkgtivnvgvvfpfppylskgtqcsvvtvdqtnvltcl 60
 QY 61 LFSGNVLEALPIIYKNGDIVFHRKLIQVYKKEGQITSSGFASLTREGTLGAPITPTSS 120
 DB 1 lfsgnvlealpiiyrkngdivfhrklioqvyykkekgtitssgsfalsltregtlgapitptss 120

DB 61 lfsngvealpiilyknodivrfhrkliqvykkekgtitssgsfalsltregtlgapitptss 120
 QY 121 KYNFTTEDIHKWVEALRWASTHMSFSTLLKLCIDYQVWQYFDLTCQLAKRAEYGCASFL 180
 DB 121 kynfttedihkwealrwasthmsfswtllklcidyqvwwqyfdltcqlakraeygcasfl 180
 QY 181 LKQWDGRTPTPPSWRVYIIDLVEGDLSHIRLQNTLITIDILVYDNIHVARSLKYGSPFLK 240
 DB 181 lkwdgtrtpppswrvyidilvegdlshirlnlntlitidilvydnihvarslkygsflr 240
 QY 241 TYSLHTRKLSMNSLQNTMLSLFPHLIGCTSYGKCIIVYLHESNDVYQLKDLISANLTAN 300
 DB 241 tyslhtklsmslqntmlslfphligctsygkciivylhesndvyqlkdlisaneltan 300
 QY 301 OHSQVYCOSERPDOSFSSGSVSLYEVRGQQLSATITITDHOVLEKPTLCAI LKOKAPQY 360
 DB 301 ohsqvycoserpdosfssgsvslyevrqqqlsatititdhoylekptlcalikokapqy 360
 QY 361 qhsdvycqsepdosfssgsvslyevrqqqlsatititdhoylekptlcalikokapqy 360
 DB 361 qhsdvycqsepdosfssgsvslyevrqqqlsatititdhoylekptlcalikokapqy 360
 QY 421 SKIWTYKNGCKRKAIVHFKNNCLILYLSNECLLLKCGTISEICLSNKPNSVYIPKSGH 480
 DB 421 skiwtynkgckrkavhfknnclilylsneclllkcgtiiseiclsnkpnsviprsg 480
 QY 481 EDLELDDLSAPFLDQCTTHHGYCTCYCT 507
 DB 481 edlelddlsapflldqctthhgyctcyct 507

RESULT 5
 AAM41911
 ID AAM41911 standard; protein: 307 AA.
 XX AC AAM41911:
 XX XX 22-OCT-2001 (first entry)
 XX DE Human polypeptide SEQ ID NO 6842.
 XX XX Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW Leukemia.
 XX OS Homo sapiens.
 XX PN WO200153312-A1.
 XX PD 26-JUL-2001.
 XX PF 26-DEC-2000; 2000WO-US34263.
 XX PR 21-JAN-2000; 2000US-0488725.
 XX PR 25-APR-2000; 2000US-0552317.
 XX PR 09-JUL-2000; 2000US-0598042.
 XX PR 19-JUL-2000; 2000US-0620312.
 XX PR 03-AUG-2000; 2000US-0653450.
 XX PR 14-SEP-2000; 2000US-0662191.
 XX PR 19-OCT-2000; 2000US-0693036.
 XX PR 29-NOV-2000; 2000US-0727344.
 XX PA (HYSE-) HYSE INC.
 XX PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 XX PI Zhao Q, Zhou P, Goodrich R, Drmanac R;
 XX DR WPI: 2001-442253/47.
 XX DR N-PSDB: AA161067.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2002, 20:29:26 ; Search time 3737 Seconds
(without alignments)
8409.721 Million cell updates/sec

Title: US-09-816-248-12

Perfect score: 1905
Sequence: 1 atgtctttgttcagcaac.....ttgcagaagatgtaatactaa 1905

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_st.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_in.*

18: em_in.*

19: em_om.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_st.*

27: em_sy.*

28: em_un.*

29: em_vl.*

30: em_htgo_hum.*

31: em_htgo_inv.*

32: em_htgo_rtd.*

33: em_htg_hum.*

34: em_htg_inv.*

35: em_htg_rtd.*

36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	1905	100.0	2631	9	AK001230	AK001230 Homo sapi
2	1901.8	99.8	2029	9	BC002923	BC002923 Homo sapi
3	1901.8	99.8	2077	9	AK001935	AK001935 Homo sapi
4	1901.8	99.8	2383	9	AK022580	AK022580 Homo sapi
5	1842.6	96.7	3192	9	AB066545	AB066545 Macaca fa
6	1268.4	66.6	1752	9	AK026234	AK026234 Homo sapi
7	408.4	21.4	2012	9	HSM800212	AL050120 Homo sapi
8	208.6	11.0	154959	9	AC004925	AC004925 Homo sapi
9	104.8	5.5	320	11	G12915	G12915 swSS2590 Er
10	51.2	2.7	1141	6	AX083744	AX083744 Sequence
11	45.8	2.4	127650	1	MPULM01	AL445563 Mycoplasma
12	45	2.4	110000	2	PFMAL13P2_1	Continuation (2 of
13	43.8	2.3	1141	6	AX083744	AX083744 Sequence
14	42.8	2.2	34641	3	DMBN32G13	AL035631 Drosophil
15	42	2.2	96310	9	AL136444	AL136444 Human DNA
16	42	2.2	159475	2	AC021378	AC021378 Homo sapi
17	42	2.2	169421	2	AC036163	AC036163 Homo sapi
18	42	2.2	172537	2	AC073122	AC073122 Homo sapi
19	42	2.2	174672	9	AC039057	AC039057 Homo sapi
20	42	2.2	182048	2	AC027439	AC027439 Homo sapi
21	42	2.2	182383	2	AC021745	AC021745 Homo sapi
22	41.6	2.2	181596	2	AC068158	AC068158 Homo sapi
23	41.4	2.2	60565	2	AC023852	AC023852 Homo sapi
24	41.4	2.2	153724	2	AC041034	AC041034 Homo sapi
25	41.4	2.2	209258	2	AC019234	AC019234 Homo sapi
26	41.2	2.2	40255	3	U41272	U41272 Caenorhabdi
27	41.2	2.2	110000	2	CEY39B6_1	Continuation (2 of
28	41.2	2.2	110000	2	CEY39B6_2	Continuation (3 of
29	41.2	2.2	154386	2	AC016314	AC016314 Homo sapi
30	41.2	2.2	158581	2	AL356485	AL356485 Homo sapi
31	41.2	2.2	180283	2	AL592406	AL592406 Homo sapi
32	41.2	2.2	207022	3	CEY39B6B	AL132896 Caenorhab
33	41.2	2.2	246237	3	CEY39B6A	AL132948 Caenorhab
34	41	2.2	7152	1	BA009185	U09185 Buchnera ap
35	40.8	2.1	39726	3	CER0411	292824 Caenorhabdi
36	40.8	2.1	150993	2	AC024289	AC024289 Homo sapi
37	40.8	2.1	195368	2	AC068399	AC068399 Homo sapi
38	40.8	2.1	198551	2	AF284562	AF284562 Homo sapi
39	40.6	2.1	124004	9	AC004947	AC004947 Homo sapi
40	40.6	2.1	181561	2	AC015911	AC015911 Homo sapi
41	40.4	2.1	2759	3	MILTMCX1	X02438 Leishmania
42	40.4	2.1	20992	3	LEKPMAX	M10126 Leishmania
43	40.4	2.1	123159	2	AC091880	AC091880 Homo sapi
44	40.4	2.1	144598	2	AC092946	AC092946 Homo sapi
45	40.4	2.1	146602	2	AC064795	AC064795 Homo sapi

ALIGNMENTS

RESULT 1

AK001230

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (sites)

Isopai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,

Nishikawa, T., Naito, K., Sugano, S., Shiratori, A., Sufo, H.,

Wagatsuma, M., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H.,

Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,

Ono, Y., Takiuchi, S., Watanabe, S., Kimura, K., Murakami, K.,

Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,

Nakamura, Y., Nagahara, K., Masuho, Y., Ninomiya, K. and Iwayanagi, T.

AK001230 2631 bp mRNA PRI 22-FEB-2000
Homo sapiens cDNA FLJ10368 fis, clone NT2RM2001544.

AK001230
AK001230.1 GI:7022354

oligo capping: fis (full insert sequence).

Homo sapiens teratocarcinoma cell line: NT2 cDNA to mRNA,

clone: lib-NT2RM2 clone: NT2RM2001544.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2631)
AUTHORS Isogai,T. and Otsuki,T.
JOURNAL Direct Submission
Submitted (16-FEB-2000) to the DDBJ/EMBL/Genbank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory: 1532-3 Yama, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute of Virology, Institute of Medical Science, University of Tokyo.
FEATURES
source
1. .2631
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NT2"
/cell_type="Teratocarcinoma"
/clone="NT2RM2001544"
/clone_11d="NT2RM2"
/note="cloning vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells."
24. .1928
/note="unnamed protein product"
/codon_start=1
/protein_id="BAA91568.1"
/db_xref="GI:7022355"
/translation="MSI V P A T N Y I T P L N L K G T I V N Y G V K F K P Y L S K G T D Y C S V T I A D O T N V K L C L F C T N Y E A L P I Y K N G D I V P H R L K I O V Y K K E P G I T S S G F A S L P E G T I G A P I I P R T S S K Y F N T E D H M M V E A L R V M A S T H M S P S M T L L K L D Y O M O Y P D L T G C I L C K A F P C A S P L L K Y M D C T R T P S K V L T O D I V E G H S I H I R L O N I T T D I L V Y D N I V N V A S L A V G S F L A I S L S L I K L S N S N S N O T L H G C T S Y C R G I K V P E S N S D V D Q L K D L E S A N L I V A N O H S D V I C S E P D S P S S G S V S L Y E V E R C O D L S A T I L D I O Y L E R T P L C A I L K O K A P O Y R I R A K L S Y K P R L F O S V K L H C P K C H L L O E P H E G D L D I I F O D G A K T P P V K L O N T S L Y D S K I W T T K M O K R K A V H R V K N N G I L P L S N E C L L I E G G T L S E I C K L S K N F S V I P V R S C H E D L E L D I S A P L I O G T I H Y K C O C S S L R S I O N L S I V D K T S M I P S S V A L C I I V I O Y V F W T P L D O C T G L E A V L M D S K F P Q I P A S E V L M D D L O K S V D M I M D F C P R G I K I D A Y P W L E C F I K S I N V N T N G T D M O I C Y I F D T T V A D V Y"
BASE COUNT 826 a 473 c 491 g 841 t
ORIGIN
Query Match 100.0%; Score 1905; DB 9; Length 2631;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1905; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 324 TTTCATCTTTGAGCTTTGAGGCAACTTTGGAGCCCTTATCATCTCGACTTCAAC 383
Oy 361 aagattttaaacttaacttaacttaaggaccacaaatgtagaaaccccttaactgtttggga 420
Db 384 AAGTATTTTAACTTCACTACTGAGGACACAAAATGGTAGAACCCCTACGCTTGGGCA 443
Oy 421 tctactcaatgtcaaccgcttcggacatlaaataatgtagatgtagcgaatgtag 480
Db 444 TCTACTCATATGTCACCGCTCTTGACATTCTAATAATGTGTGATGTTCAGCAATTCAG 503
Oy 481 tatttgaactgacttgtagccttcctgggcaaaagagagtagagagacatctct 540
Db 504 TATTTTCACTGACTTGTGAGCTTGTGGCAAGCAAAATGGACGACCATATTTCTT 563
Oy 541 ctlaaaagtatggatagcagcagacacatctccatcttgagagctctaaacaaagac 600
Db 564 CTAAACGATATGGCATGGCACACGACACCATTTTCATCTTCCAGACTTTATACAGAC 623
Oy 601 ctgttctcttgaaagtagatlllaagatgtagatccatcgccttaacaaatct 660
Db 624 CTCTCTCTTCAACCTGATGATTAACACATCCATCGGCTCAAAATCTGCAATTAACAT 683
Oy 661 ttagctacgaataacgaatgtctcagtgagcaagatctctgaaggttgaaagcttctga 720
Db 684 TTAGCTTACGATTAACCATGTTTCATGTGGCAAGATCTCTGACCTTTCGAGCTTCTTGA 743
Oy 721 atctatagccttcaatcaaacacttcaatcaatgaaatgaaatcagaagatgaaat 780
Db 744 ATCTATTAACCTTCAATCAAAACCTTCAATCAATCAATCAATCAATCAATCAATCAAT 803
Oy 781 ttgaagttcatalctcatalgaaggtacccagttacggttcgagagatcagagcttcgcaaa 840
Db 804 TTGACTTTTCACTTTCATGAGAGTACCACTTACGCTCGGGAATGACGCTTCCACAA 863
Oy 841 aatlaactctgaatgtagatcaactgaaatgaaatgaaatgaaatgaaatgaaatgaaat 900
Db 864 ACTAACTCATCTGATCACTGATCACTGATCACTGATCACTGATCACTGATCACTGATCACT 923
Oy 901 cagacatcagagatgtagctcagatcagatcagatcagatcagatcagatcagatcagat 960
Db 924 CAGCATTCACATGTTATCTGCTCATTCATTCAGAACCTGACGACAGCTTTCAGACCTGCAATCA 983
Oy 961 gtagcatatcagaggttagaaatgtagatcagatcagatcagatcagatcagatcagatcagat 1020
Db 984 GTATCATTTATACAGAGTACAAAGATGTCAAAGCTATCTGCTACATTAACAGATCAT 1043
Oy 1021 cagatattgagagagagacacactatgtagcatcttgaaacaaagaagctcccaacatag 1080
Db 1044 CAGTATTTTGGAGAGGACACCACTATGTGCAATTTTGAACAAAGCTCTTCACAAATAC 1103
Oy 1081 cgcattcgaagcaaaatgtagatcatalaagccagaagactatctcagctctgtaaact 1140
Db 1104 CCCATCCGAGCAAAATTTGAGGTCAATTAAGCCCAAGAACATTTTCAGCTGTTAAACTT 1163
Oy 1141 catctccctaaatgtagatctcagatcagatcagatcagatcagatcagatcagatcagat 1200
Db 1164 CATTCGCCCTAAATGTCAATTTGCTGCAAGAACTTCCACATGAGGGGATTTGATTAAT 1223
Oy 1201 ttcagagatgtagcaactaaaaaccccaatgtagatcagatcagatcagatcagatcagat 1260
Db 1224 TTTCAGAGATGTGCAAACTAAACCCAGTTGCAAGTTTACAAATATCATTTATATGAT 1283
Oy 1261 tcaaaatctgtagcacaatcuaaaatcuaaaatgtagaagtagagatcagatcagatcagat 1320
Db 1284 TCAAAATCTTGACACATTAATAATCAAAAGACAAAGATGACAGTTCAATTTCTGATAA 1343
Oy 1321 aatlaatgtagatctccgcttccaaatgaaatgtagatcagatcagatcagatcagatcagat 1380
Db 1344 AATATATGATTTCTCCCTTTCAATATGATGTCTACTTTGATAGAAAGAGATCACTC 1403
Oy 1381 agtgaatctgcaaacctcctggaagaagatlaaagtagtaactctcgttgagatctgagcac 1440
Db 1404 ACTGAATTTGCAAACTGTCGAAACGTTTAATGCTTAATTTCTGCTGACATCTGCCCCAC 1463

Oy 541 ctcaaggatgagatgacacgacacacattccattcttgagagatcttaacacagac 600
 Db 607 cttaaggatgagatgacacacacacacacacacacacacacacacacacacac 666
 Oy 601 ctgctctgagagatgagatgagatgagatgagatgagatgagatgagatgagat 660
 Db 667 ctgctctgagagatgagatgagatgagatgagatgagatgagatgagatgagat 726
 Oy 661 ttagctcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 720
 Db 727 ttgagctcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 786
 Oy 721 atcctcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 780
 Db 787 atctcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 846
 Oy 781 ttagagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 840
 Db 847 ttgagctcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 906
 Oy 841 agtcaactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 900
 Db 907 agtcaactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 966
 Oy 901 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 960
 Db 967 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1026
 Oy 961 gtagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1020
 Db 1027 gtagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1086
 Oy 1021 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1080
 Db 1087 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1146
 Oy 1081 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1140
 Db 1147 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1206
 Oy 1141 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1200
 Db 1207 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1266
 Oy 1201 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1260
 Db 1267 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1326
 Oy 1261 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1320
 Db 1327 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1386
 Oy 1321 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1380
 Db 1387 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1446
 Oy 1381 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1440
 Db 1447 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1506
 Oy 1441 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1500
 Db 1507 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1566
 Oy 1501 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1560
 Db 1567 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1626
 Oy 1561 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1620
 Db 1627 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1686
 Oy 1621 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1680

Db 1687 GTGTTTGTATGACCTTTACACTTGATGATGAGAACACAGACTACTAGACCTATTCATG 1746
 Oy 1681 gattctgacaaattcttccacacacacacacacacacacacacacacacacacacac 1740
 Db 1747 gattctgacaaattcttccacacacacacacacacacacacacacacacacacacac 1806
 Oy 1741 aaaaagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1800
 Db 1807 aaaaagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1866
 Oy 1801 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1860
 Db 1867 cagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1926
 Oy 1861 ttagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1905
 Db 1927 ttagcactcagatgagatgagatgagatgagatgagatgagatgagatgagatgagat 1971

RESULT 3
 AK001935 2077 bp mRNA PPI 22-FEB-2000
 LOCUS Homo sapiens cDNA FLJ11073 f1s, clone PLACE1005026.
 DEFINITION AK001935
 ACCESSION AK001935
 VERSION AK001935.1 G1:7023508
 KEYWORDS oligo capping; f1s (full insert sequence).
 SOURCE Homo sapiens placenta cDNA to mRNA, clone_11b:PLACE1
 clone:PLACE1005026.

ORGANISM

Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (sales)
 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nishii,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ishii,S., Kawaj,Y., Saito,K., Yamamoto,J., Wakematsu,A., Nakamura,Y., Nagahori,K., Masuno,Y. and Sasaki,N.

AUTHORS

NEDO human cDNA sequencing project
 Unpublished (2000)
 2 (bases: 1 to 2077)
 Isogai,T. and Otsuki,T.

TITLE

NEDO human cDNA sequencing project
 Submitted (16-FEB-2000) to the DDBJ/EMBL/Genbank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yuna, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)

JOURNAL

NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction; 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute of (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

COMMENT

NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction; 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute of (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.

FEATURES

Location/Qualifiers
 1..2077
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="PLACE1005026"
 /clone_11b="PLACE1"
 /issue_type="Placenta"
 /note="Cloning vector: pME18SF3"
 131..2035
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAA91988.1"
 /db_xref="GI:7023509"
 /translation="MSLVPAHYIYTPINOLKGTIVNYGVYKFFKPYLSKTDYCSVTITVDQNVRLTCLPFSNGEALPIYKGDVFRHRLKIOYTKETGITSSTGFA SLTPEGTGAPLIPTSSKYFNFTEDHKWEALRWASTHMSFWTLKLCVOPKO YFDLCOILGKAELVGSFLKVMGDTPTPPSNRVLIODLVLEGLDLSHRLQNLIT I DLYVDNHVARSILKVSFLRITSLATKLOSMSNEOTMLSEFLHSGTSGRGRI VLPENSVDOLKRLDESANLITJANOHSYVTCSPSPSSGVSLEYERCOOLSA TLTIDHVLKPLCALIKOKAPVDYRIKALRSYKRRRLPQSVKLCRCHLQLEVP

source

CDS
 131..2035
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAA91988.1"
 /db_xref="GI:7023509"
 /translation="MSLVPAHYIYTPINOLKGTIVNYGVYKFFKPYLSKTDYCSVTITVDQNVRLTCLPFSNGEALPIYKGDVFRHRLKIOYTKETGITSSTGFA SLTPEGTGAPLIPTSSKYFNFTEDHKWEALRWASTHMSFWTLKLCVOPKO YFDLCOILGKAELVGSFLKVMGDTPTPPSNRVLIODLVLEGLDLSHRLQNLIT I DLYVDNHVARSILKVSFLRITSLATKLOSMSNEOTMLSEFLHSGTSGRGRI VLPENSVDOLKRLDESANLITJANOHSYVTCSPSPSSGVSLEYERCOOLSA TLTIDHVLKPLCALIKOKAPVDYRIKALRSYKRRRLPQSVKLCRCHLQLEVP

HEGDLIIIFODGATKTPDVVKLQNTSLYDSKIWTTKNQKGRKVAHFVKNNGILPLSNE
CLLIEGGTISEIKLSNKFNSVPIVRSHEDLELDLSAPFLIOGTIIHYGCKQCSS
LRSTONLSLVDTKSWIPSSVAELGIVPQYVFVMTFTLLDDCTGVLEAYLMDSDKFF
QIPASEVMDODDLQKSDVMDIMDFCPGPKIDAYPWLECFIKSYNTNGTDNOICVQI
FDTTVASBVI*

BASE COUNT 657 a 401 c 393 g 626 t
ORIGIN

Query Match		99.8%	Score 1901.8	DB 9	Length 2077		
Best Local Similarity		99.9%	Pred. No. 0				
Matches 1903		Conservative 0	Mismatches 2	Indels 0	Gaps 0		
QY	1	atgtcttttgggtccagcaacaattatatacacacccctgaatcaacttaagggtggt	60				
Db	131	ATGTCTTTGTTCCAGCAACAATATATATACACCCCTGAATCAACTTAAAGGGTGT	190				
QY	61	acaattgcaatgtctatggtgtggaagtcttttaagcccccatatctaagcaagga	120				
Db	191	ACAAATTGTCATGTCTATGGTGTGTGGAAGTCTTTAAGCCCCCATATCTAAGCAAGGA	250				
QY	121	actgattatgtctcagttgtaactattgtgacagacaatgttaaaactaaacttgccgt	180				
Db	251	ACTGATTATGTCTCAGTTGAATATTGTGGACCCAGACAAATGTAAACTTAACTTGCCTG	310				
QY	181	ctcttttagtgaactatgaagcccttccataatttataaaatggaatattgttcgc	240				
Db	311	CTCTTTTACGTGGAACATATCAAGCCCTTCCAAATAATTTATAAAATGGAGATATTGTTGCGC	370				
QY	241	ttccacagctgaagatccaagtatataaaaggagactcaggtatataccacagcttcggt	300				
Db	371	TTTCACAGCGCTGAAGATTCAGATATATAAAGGAGACTCAGGGTATCACCAGCTCTGGC	430				
QY	301	tttgcatctttgacgtttgaggaactttgggaagccctatcatatcgtcagcttcaagc	360				
Db	431	TTTTCATCTTTGACGCTTTGAUGGAACCTTTGGGAGCCCTATCATACCTCGCACTTCAAGC	490				
QY	361	aagtattttaaacttcaactactgaagaccacaaatgtagagcttcaggttcttgagga	420				
Db	491	AAGTATTTTAACTTCACTACTAGGAGCCCAAAATGGTAGAAGCTTACGTTGTTGGCA	550				
QY	421	tctactcatatgcaacgcttcttgacatatactaaaattgtgtgatttcaagccaatgcag	480				
Db	551	TCTACTCATATGTCACGCTCTTGGACATTATCAAAATGTGTGATGTTCAAGCAATGCAG	610				
QY	481	tattttgacgtgactgttcagctctctgggaagcaaaatgtagagcagagctattctct	540				
Db	611	TATTTTGACCTGACTTTGTGACGCTTTGGGCAAAAGCAGAAAGTGGAGCAATCATTTCTT	670				
QY	541	ctaaaggtatggtgatggtcaccaggaacacatttccattttgagaggtctttaaatacagac	600				
Db	671	CTAAAGGTATGGGATGGCAGCAGGACACCAATTTCCATCTTGCAGAGTCTTTAATACAGAC	710				
QY	601	cttgctcttgaaggtgatttaagtcaactcaactcgggtcacaataatctgacaataagacatt	660				
Db	731	CTTGTTCTTGAAGGTGANTTAAAGTCACATCCATCGGCTTACAAAATCTGCACAAATAGACATT	790				
QY	661	ttagtctcaagaataacatatttcatgtggcaagatctcttgaaggtcttgaaggtcttctaga	720				
Db	791	TTAGTCTACGATAAKCATGTTCTCATGTGGCAGATCTCTGAAAGTTGGAGGCTTCTTTAGA	850				
QY	721	atctatagctctcaataccaaacttcaatcaatgaattcagaagaatcagaagaatttaagt	780				
Db	851	ATCATATAGGCTTTCATACCAAACTTCAATCAATGAATTCAGAAATCAGACAAATGTAAAGT	910				
QY	781	ttaagatttcaactctcaggtgaggtaccagtttacggtcgggaatcaaggtcttgccagaa	840				
Db	911	TTAGAGTTCATCTTCATCGAGCTACCAAGTTACGGTCGGGGAATCAGGGTCTTCCAGAA	970				
QY	841	agtaactctgattgttgatcaactgaaaaaggaatttagaattctgcaaatcttgacagccaat	900				
Db	971	AGTAACCTCTGATGTGATCACTGAAAAAGGATTTAGAAATCTGCAAAATTTTGACAGCCAAAT	1030				

RESULT 4
AK022580
LOCUS
DEFINITION

2383 bp mRNA
Homo sapiens cDNA FLJ12518 fis, clone NT2RM2001805.

29-SEP-2000
PRI

QY	901	cagcattcagatgtattctctgtcaatcagaacctgacgacagctttccaaagctctggatca	960
Db	1031	CAGCATTTCAGATGTTATCTGCTCAATCAGAAACCTGACGACAGCTTTTCCAAGCTTGGATCA	1090
QY	961	gtatcattatcagaggtagaagaatgtcaacagctatctgctacaatacttacagatcat	1020
Db	1091	GTATCATTTATACGAGGTAGAAAGATGTCAACAGCTATCTGCTACAATACTTACAGATCAT	1150
QY	1021	caattattggagagagaccactatattgcatcttttaaaacaaaaagctcctcaacaatac	1080
Db	1151	CAGTATTTGGAGAGGACACACCTATCTGCTCAATTTTCAAAACAAAAGCTCTCTCAACAATAC	1210
QY	1081	cgaatcagagcaaaatbgaggttcataataagccacagaagactatttcagctctgttaaacct	1140
Db	1211	CGCATCCGAGCAAAATTTGAGTTCATATAAGCCACAGAACTATTTTCAGTCTGTGTTAACTT	1270
QY	1141	catgcccataatgtcatcttgctgcaagagttccacatagaggcgatttggataaatt	1200
Db	1271	CATTGCCCTAAATGTCATTTCTTGCAGAAAGTTTCCACATGAGGGCGATTGGATATAATT	1330
QY	1201	tttcaggatgggtgcaactaaacccagttgtcaagttacaaaaatcacatcattatcat	1260
Db	1331	TTTTCAGGATGCTGCAACTAAACCCGAGATGTCAGCTACAAAATACATCATTTATATGAT	1390
QY	1261	tcaaaaatctggaccactaaaaatcaaaagacgaaaagtagcagttcatctttgtgaaa	1320
Db	1391	TCAAAATCTGGACCACCTAAAAATCAAAAGGACGAAAGACTAGCAGTTCTATTTGTGAAA	1450
QY	1321	ataaattggtatctctcggctttcaaatgaaatgctactcttttgatagaagaggtacactc	1380
Db	1451	AATAATGGTATTCTCCGGCTTTCAAAATGAATGTCTACTTTTGATAGAAAGAGGTAGACTC	1510
QY	1381	agtgaaatttgcaaacctctgcaacaagtctaaatgtagtgaattctctgtgagatctggccac	1440
Db	1511	AGTGAATTTGCAAACTCTCGAAACAAGTTTAAATGCTAAATTCCTGTGAGATCTGGCCAC	1570
QY	1441	gaagacctgaaacttttggacctttcagacccaattcttatacaaaagaaacaatcacatcac	1500
Db	1571	GAAGACCTGCAACTTTTGGACCTTTTCAGCACCACTTTCTTATACAGGAACAATACATCAC	1630
QY	1501	tacggatgtaaacagtgctctatgttctgagatccatcaaaaaataatccctcgtgtgat	1560
Db	1631	TATGGATGTAACAGAGTTCCTAGTTTCAGATCCATCCAGATTCGAAATTAANTTCCCTGGTTGAT	1690
QY	1561	aaaacatctggaqctctctctctgtgaggaagcactgggtatctgtacacctccaatat	1620
Db	1691	AAAACATCGTGGATTCCTTCTCTGTCGAGAAAGCACCTGGGTATTCTACCCCTCCAATAT	1750
QY	1621	gtcttcttctatgacctttcacacttgatggaacagggagctactagaagcctatctcatg	1680
Db	1751	GTCTTTGTTATGACCTTTTACACTTGTATGATGGAACAGGAGTACTAGAACCTTATCTCATG	1810
QY	1681	gattctgacaanaattcttccagatctcagagcatcagaaggtcttgaatggatgagaccttcag	1740
Db	1811	CATTCTGCACAAATTTCTCCAGATTCAGATCCAGATCAGAAAGTTCTCATGATGATGACCTTCAG	1870
QY	1741	aaaatgttgaatctgatactatgataatcttctctcctccaggaataaaaaatctgatacat	1800
Db	1871	AAAAGTGTGGATATGATCATGATATGTTTTTGTCTCCAGGAATAAAAAATTTGATGGATAT	1930
QY	1801	ccgtgggtgaaatgcttctcacaagtcatacaaaatggaacagataatcaaat	1860
Db	1931	CCGTGGTTGGAATGCTTTCATCAAGTATCAGATGTCACAAATGGAACAGACATAATCAAAAT	1990
QY	1861	tgctatcagatcttttgacaccacagttggagaagatgtaatactcaa	1905
Db	1991	TGCTATCAGATTTTTCAGACCCACAGTTTGCAGAAAGATGTAATCTAA	2035

ACCESSION AK022580.1 GI:10434050
 VERSION AK022580.1 GI:10434050
 KEYWORDS Oligo cloning; fis (full) insert sequence;
 SOURCE Homo sapiens testicular carcinoma cell line; NT2 CDNA to mRNA,
 clone_11b:NT2RM2 clone:NT2RM2001805.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (sites)
 AUTHORS Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
 Nishikawa, T., Nagai, K., Sugano, S., Aotsuka, S., Yoshikawa, Y.,
 Matsunawa, H., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J.,
 Wakamatsu, A., Nakamura, Y., Nagahari, K., Masuko, Y. and Sasaki, N.
 NEDO human cDNA sequencing project
 TITLE Unpublished (2000)
 JOURNAL 2 (bases 1 to 2383)
 REFERENCE 1 Isogai, T. and Otsuki, T.
 AUTHORS Direct Submission
 TITLE Submitted (23-AUG-2000) to the DDBJ/EMBL/Genbank databases. Takao
 Isogai, Helix Research Institute, Genomics Laboratory: 1532-3 Yama,
 Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp,
 Tel:81-438-52-3951, Fax:81-438-52-3952)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan: cDNA full insert
 sequencing. Research Association for Biotechnology: cDNA library
 construction, 5' & 3'-end one pass sequencing and clone selection:
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.
 FEATURES
 source Location/Qualifiers
 1..2383
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone="NT2RM2001805"
 /clone_11b="NT2RM2"
 /note="cloning vector: PMEB8FL3-mRNA from uninduced NT2
 neuronal precursor cells."
 439..2343
 /note="unlabeled protein product"
 /codon_start=1
 /protein_id="BAB14110.1"
 /db_xref="gi:10434051"
 /translation="MSLVPAIVYITPLNOLKGGTVVNVGVKFKPPYLSKGTDC
 SVYTVIDQTNVTKICLLPNSGNELPIIKMGDIYRFRRLKIOYKKEQSTSSGA
 SLTTEGTGAPRIIPRTSSKYPNFTDEHKWELRYVASTHMSMTLLKLDYOPMO
 YFDLTCOLLGAKEVDGASFLKVMGDTTPPSRNVLLQDLYEGDLSHTRDLONTI
 DLYVDNHVHARSLKVSFLKITSYLRKLDKMSSENOTMLSEPHLHGISTYGRIR
 VLPESNSVDOLKDLKESANLTANOHSVITQSESDSPSSGSVLYEVCRCOOLSA
 TILTHOYLERTPLCAILKOKAPOYRIIRAKLRKYPKRRLLQSVKLNHC PKHLLQEV
 HEGDLDIIFQDGATKTPVKYKONTSLYDSKIMTWTNOKGRVAHFVNNNLLPLSNE
 CLLTEGGLISFICKLSPNSVLPVRSKHEDLELLDSAPFLLOGTTHRGCCGCS
 LRSIONLSLVDTKSMIPSSVAEALGTYPLQYVFMPTTLDGSGVEAYLMDSKFP
 QIPASEVLMDDDLQKSDVMIMDMFCPPKIKIDATPMLCEFTKSYNVNNGTNOICYOI
 FPTTYAEVPI"
 BASE COUNT 750 a 456 c 459 g 718 t
 ORIGIN
 Query Match 99.8%; Score 1901.8; DH 9; Length 2383;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1903; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 559 ACTGATTATGCTCAGCTGTGACTATTTGTGGACGACGAAATGTAAACTTACTTGCTG 618
 OY cttcttaagtggaactatgaagcccttcacaacttataaaatlgagatattgctgc 240
 Db 181 cttcttaagtggaactatgaagcccttcacaacttataaaatlgagatattgctgc 240
 Db 619 cttcttaagtggaactatgaagcccttcacaacttataaaatlgagatattgctgc 678
 OY 241 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagc 300
 Db 679 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagc 738
 OY 301 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagc 360
 Db 739 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagc 798
 OY 361 aagatatttaactcgaagatcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc 420
 Db 799 aagatatttaactcgaagatcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc 858
 OY 421 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 480
 Db 859 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 918
 OY 481 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 540
 Db 919 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 978
 OY 541 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 600
 Db 979 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1038
 OY 601 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 660
 Db 1039 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1098
 OY 661 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 720
 Db 1099 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1158
 OY 721 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 780
 Db 1159 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1218
 OY 781 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 840
 Db 1219 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1278
 OY 841 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 900
 Db 1279 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1338
 OY 901 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 960
 Db 1339 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1398
 OY 961 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1020
 Db 1399 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1458
 OY 1021 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1080
 Db 1459 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1518
 OY 1081 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1140
 Db 1519 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1578
 OY 1141 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1200
 Db 1579 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1638
 OY 1201 tttcagaagcgaagatcgaagatataaaaggaagcgaagatcgaagcgaagcgaagcgaagc 1260

Db	1639	TTTCAGGATGTTGCACTAAACCCAGATGTCAGCTACAAATACATCATATATGAT	1698
QY	1261	tcataaaatctgaccactaaaatacaaaagacgaagtagcaggttcattttgtgaaa	1320
Db	1699	TCAAAATCTGACCACATAAAATCAAAAAGGACGAAGTAGCAGTTCAATTTGTGAAA	1758
QY	1321	aataatggtattctcccgctttcaaatgaaatgctactctttgatagagaggtacactc	1380
Db	1759	AATAATGTAATTCGCCGCTTTCAATCAATGCTCTACTTTTGATAGAGGAGGTACACTC	1818
QY	1381	agtgaatttcacaaactcgaacaagttaaataagtgtaattcctctgagatctggccac	1440
Db	1819	AGTGAATTTGCAAACTCGAACAAGTTTAAATAGTGAATTTCTCTGTGAGATCTGGCCAC	1878
QY	1441	gaagacctggaaacttttgaccttcagaccatttcagaccatttcctatacaagaacaatacatcac	1500
Db	1879	GAAGACCTGGAACTTTTGACCTTTTCAGCCCTTCAGCCACTTTCTTATACAGGAACAATACATCAC	1938
QY	1501	tatgatgtaaacagtggtcttagttgagatccatacaaaatcctaattccctgggtgat	1560
Db	1939	TATGATGTAACAGTGTTCTAGTTTCAGATCCATACAAATCTAAATTCCTGTTGAT	1998
QY	1561	aaacatcgtgattcctctctctgctggcgaagcactgggtattgtaccctcccaatat	1620
Db	1999	AAACATCTGTTGATTCCTTCTCTGTCGCAAGCACTGGGTATTGTACCCCTCCAATAT	2058
QY	1621	gtgtttgtatgacctttcacacttgatggaaacagagtagtactagaagcctatctcatg	1680
Db	2059	GTGTTGTTATGACCTTTACACTTGATGATGGAACAGGACTACTAGAGCCTATCTCATG	2118
QY	1681	gattctgacaaattcttccagatccagatccagaagttctgtatgatgatgacccctcag	1740
Db	2119	CATTCTGACAAATTTCTCCAGATTTCCAGATTCAGACAGTCTGATGATGATGATGATGAT	2178
QY	1741	aaaagtggtgatgatgatgatgatgttttctctccagaggaataaaattcattgat	1800
Db	2179	ANAGTGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2238
QY	1801	ccgtggtggaatgcttccatcaagtcacaaatgacaaatgacaaatgacaaatgacaaat	1860
Db	2239	CCGTGTTGGAATGCTTCATCAAGTCATACATGTCACAAATGGAACAGATGATGATGAT	2298
QY	1861	tgcatacagatttttgacacacacagttgcgaagatgataactca	1905
Db	2299	TGCTATCAGATTTTTCACACCACAGTTTCCAGAGATGTAATCTAA	2343
RESULT	5		
LOCUS	AB066545	3192 bp	mRNA
DEFINITION	Macaca fascicularis brain cDNA clone:QtrA-10940, full insert		
ACCESSION	AB066545		
VERSION	AB066545.1	GI:15021899	
KEYWORDS	oligo capping; fls (full insert sequence).		
SOURCE	Macaca fascicularis adult male temporal lobe right cDNA to mRNA, clone:lib:macaque brain cDNA library QtrA clone:QtrA-10940.		
ORGANISM	Macaca fascicularis		
REFERENCE	1 (bases 1 to 3192)		
AUTHORS	Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M., Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.		
TITLE	Isolation of full-length cDNA clones from macaque brain cDNA libraries		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 3192)		
AUTHORS	Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUL-2001) Katsuyuki Hashimoto, National Institute of Infectious Diseases, Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan		

(E-mail:khashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/, Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)	
Lab host:	TOP10
Vector:	pME18S-FL3 (Acc.No. AB009864)
R. Site1:	DraIII (CACTGTGTG)
R. Site2:	DraIII (CAGCATGTG)
Description:	1st strand cDNA was primed with an oligo(dT) primer [ATCTGCGCTTTTCTTTTCTTTT], double-stranded cDNA was synthesized using specific 5' and 3' primers and amplified by PCR. The PCR product was digested with SfiI and size selection was performed to exclude fragments <1.5kb. The SfiI-digested PCR product was cloned into distinct DraIII sites of pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Libraries were constructed by Sugano et al.(University of Tokyo, Institute of Medical Science). Custom primer used for sequencing
(5' end primer [CTTCTGCTCTAATAACCTGCG];	
3' end primer [CGACCTCCACCTCGACACA]);	
Location/Qualifiers	
1. .3192	
/organism="Macaca fascicularis"	
/db_xref="taxon:9541"	
/clone="QtrA-10940"	
/sex="male"	
/tissue_type="temporal lobe right"	
/clone_lib="macaque brain cDNA library QtrA"	
/dev_stage="adult"	
569. -2473	
/codon_start=1	
/product="hypothetical protein"	
/protein_id="BAB62219.1"	
/db_xref="GI:15021899"	
/translation="MSLVPATNYIYPLNQLKGGTIVNVYGVVKKFPKPPYLSKGTGYC	
SVWTVDGTRVETLLFSGNYEALPIYKNGDIVRFHRLKIOYVKKTOGTSSGFA	
SITFESTLGAPIIPRTSSKYFNFTEDHKMVELRWASTHMSPSDWLTKLCDVOPMO	
YFDLTCOLLGKRAEVDGAFLLKVPDTRTPPSRWLVLDLVLEGSWLHSHIRLQNLTI	
DILVTDNHVHVARSKVGSFLRIYSHTKLQSMNSNQTMLSLEPHLHGGSYGRGIR	
VLPTNSDVQQLKDLKESANTANHISDVICQSEPDSPSSGVSLSLYEVEKCOQLS	
TILTDHQLERTPLCALIKAKAPQOYIRAKLSYKPRFLQSVKHLGCHLQLEVP	
REGDLDLIDQAGTKPDVKLQNTALYDSKLTWTKNQRKRVAVHFNKNGTLPLSNE	
CLLLLEGTLSEICKLSNKNVSIPIVRSCHDELLELDSAPFLQCTTHHYCKQCSS	
LRSTONLSLVDKTSKIPSSVAEVLGIVPLQYVFMVFTLDDGTGVLAYLMDSDKPF	
QIPASEVLMDLULQKSDMIMDMFPPGKIDAYIPWLECFIKSYNVTNLTQYCIQI	
PDITVAEDVI"	
BASE COUNT	1005 a 564 c 596 q 1027 t
ORIGIN	
Query Match	96.7%; Score 1842.6; DB 9; Length 3192;
Best Local Similarity	98.0%; Pred. No. 0;
Matches 1866; Conservative	0; Mismatches 39; Indels 0; Gaps 0;
QY	1 atgtcttttgcacacacaaattatataatacacccctgaatcaacttaagggtggt 60
Db	569 ATGCTCTTTGTTCCAGCACAAATTATATATATATACACCCCTGATCACTTAAGCGTGT 628
QY	61 acaatgtcaatgtctatgtgttbtggaagttctttaagcccccacatatataaagcaaaaga 120
Db	629 ACAATTTGTCMAATGCTCTACCGTCTGTGAAAGTCTTTTAAGCCCCCATATCTAAGCAAGA 688
QY	121 actaattatttactcaattgttaactattgtgagaccagacaaataataaactaacttgcctg 180
Db	689 ACTGATATTATGCTAGTTGTAACAATTTGGGACCCAGACAAATGTAAATTTAACTTGCCTG 748
QY	181 cctcttagtgaactatagaagcccttccaaataattataaaaaaggagatatttcttcg 240
Db	749 CTCCTTAGTGGAAACTATGACCCCTTCCAATAATTTATAAATGGAGATATTCTTCGC 808
QY	241 ttccacaggtgaagattcaagtataataaaggagactcaggttatcaccagctctctggc 300
Db	809 TTTCACAGCTCAACATCAACTATATAAAAAAGGAGACTCAGGGTATCACCAGCTCTGCC 868
QY	301 ttgcacatttgcagcttgaggaactttggagacccttatcatatcctgcagacttcaagc 360

Db 869 TTTCGATCTTTGACGTTTGAGGCACTTTGGAGCCCTTATCATACCTCGCACTTCAAGC 928
 Oy 361 aagatcttcaactcaactcagagagacacaaatggtaagacccttaagcttgggagc 420
 Db 929 AAGTATTTTAACTTCACCTACAGAGACCAAAATGGTACAAACTTACGTGTTGGGCA 988
 Oy 421 tctaactatctgacccgctcttgagacatctacaaatctgtgagctcagccaaagcag 480
 Db 989 TCTACTCATATGTCCCGCTTGTGACATTAATAATTTGCTGTGATGGCCCAATGACAG 1048
 Oy 481 taattgaccgactgtgcaactcttgaggacaaagcagaagtgagagagacacatcttclt 540
 Db 1043 TATTTTTCGACCTTCATCTGTACACTCTGTGGCAAAACAGAACGTGCATGCCATCTTCTT 1108
 Oy 541 ctaagagatctgagagagacacagacacacatcttccatcttgagagagcttcaatcagac 600
 Db 1109 CTAAAGTATGGGATGGGACCCAGCACCATTTTCATCTTGTGACAGCTTGTATTCAAACAC 1168
 Oy 601 ctgtctctgagagtgatcttaagctcagacacacacacacacacacacacacacacacac 660
 Db 1169 CTGTGCTTGAAGGTGATTTAAGTCATCCATCGGCTACAAATCTGACAAATGACATTT 1228
 Oy 661 ttagctcaagaaacacatctgacatctgagagagacacacacacacacacacacacacac 720
 Db 1229 TTAGCTTACATTAACCATCTTCTGAGATCTCTGAAGCTTGGAAAGCTTCTTGA 1288
 Oy 721 aactatagccctcacaac 780
 Db 1289 ATCTATAGCTTCTATACCAACCTTCAATCATGATTCAGACATTCAGACATGTTAAGT 1348
 Oy 781 ttagagcttcaactcttcaagagagacacacacacacacacacacacacacacacacacac 840
 Db 1349 TTAGAGCTTCACTTCACTGAGAGTCACTATGTGTGGGCAATGAGGCTTGGCAGAA 1408
 Oy 841 aactaactcagatctgagac 900
 Db 1409 ACTTACTCTGATGTGATTAAGTCACTGAGAAAGATTTAAGATTCGCAAAATTTGCACCCAA 1468
 Oy 901 cagcatctcagatctgac 960
 Db 1469 CACCATTCACATGTATCTCTCAATCAGAACCTGATGACAGCTTCTCCAAAGCTCTGATCA 1528
 Oy 961 gttctatctatagagagtagaagaatgtaacacacacacacacacacacacacacacacac 1020
 Db 1529 GTATCATTTGATAGGTAGAAAGATCTCAACAGCTATCTGCTCAATTTACATGATCAT 1588
 Oy 1021 cagatcttgagagagac 1080
 Db 1589 CAATATTTTGGAGAGACACACACTGTGTGCATTTTGAACAAAGCTCTCTCAACATAT 1648
 Oy 1081 cagatctcagagac 1140
 Db 1649 CGCATTCGAGCAAAATTTGAGGTATCAATTAAGCCCAAGACTATTTTCAAGTCTGTAATCT 1708
 Oy 1141 catgcccataatgacatctgtcagagagacacacacacacacacacacacacacacacacac 1200
 Db 1709 CATTTGCCCTAAATGTCAATTTAGCAAGAGTTCCAGTGAAGGTGATTTGGATATTAAT 1768
 Oy 1201 ttcagagatctgac 1260
 Db 1769 TTGAGGATGTGTGCACTTAAACCCCGCATGTCAAGCTTCAAAATACAGATTTATAGAT 1828
 Oy 1261 tcaaaaatctgagac 1320
 Db 1829 TCAAAAATCTGACACACTAAAAAATCAGAAAGACGAAAGATGACAGTTCTTTGTGAAA 1888
 Oy 1321 aataatgtatctcccgcttcaaaatgagatgctactcttgatagagagaggtacacac 1380
 Db 1889 AATATGTATTTCTCCCACTTCAAAATGATGTCTACTTTGATAGAAAGAGTACACTC 1948
 Oy 1381 agtgaatctgac 1440
 Db 1949 AGTGAATTTTGCAAACCTGTCAACAGTTTAAATGTAAATTTCTGTGACATCTGGCCAC 2008

Oy 1441 gaagacccgacactcttgagacactcttcagacacacacacacacacacacacacacacacac 1500
 Db 2009 GAAGACCTGGAACTTTTGGATCTTTCAGACACAAATTTCTTATACAGAACATATACATCAC 2068
 Oy 1501 tatgagatgaacagagcttctgacttgagatcacaataatcaatctcccttgatgacat 1560
 Db 2069 TATGATGTAAAGCAGTGTCTGATTTTGAATCTTACAAATCTTAAATTTCTGCTGAT 2128
 Oy 1561 aaaaatctgagac 1620
 Db 2129 AATACATCTGTGATTTCTCTTCTGTGACAAAGTACCTGCTATTTGACCCCTCAATAT 2188
 Oy 1621 gttctgtatgacactcttcaac 1680
 Db 2189 GTGTTCTTATGACCTTTTACCTTGTATGATGAAACAGAGACTACAAAGCTTATCTCAT 2248
 Oy 1681 gttctgtatgacactcttcaac 1740
 Db 2249 GATTCGACAAATTTCTTCTGATCTCCACATCCACATCAGAACTTCAATGATGACCTTCAG 2308
 Oy 1741 aaaaatgagatgac 1800
 Db 2309 AAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2368
 Oy 1801 cctgtcttgagatgac 1860
 Db 2369 CCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2428
 Oy 1861 tgcatacagatcttgaac 1905
 Db 2429 TGCATACATTTTTCACACACACAGTTCACAGAAATGATTAATCTAA 2473

RESULT 6
 AK026234 1752 bp mRNA PRI 29-SEP-2000
 LOCUS Homo sapiens CDNA: FLJ22581 fls, clone HS102573.
 DEFINITION AK026234
 ACCESSION AK026234.1 GI:10439019
 VERSION AK026234.1 GI:10439019
 KEYWORDS cDNA; full insert sequence.
 SOURCE Homo sapiens human small intestine CDNA to mRNA, clone HS102573.
 clone:HS102573.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (sites)
 Matanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Oca,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
 NEDO human cDNA sequencing project
 Unpublished (2000)
 2 (bases 1 to 1752)
 Sugano,S., Suzuki,Y., Oca,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
 Direct Submission
 Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio International Trade and Industry of Japan: CDNA full insert sequencing: Research Association for Biotechnology: CDNA library construction, 5'- and 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

COMMENT
 NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: CDNA full insert sequencing: Research Association for Biotechnology: CDNA library construction, 5'- and 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES
 source Location/Qualifiers
 1..1752
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

```

/clone="HS102573"
/clone_lib="HS1"
/tissue_type="human small intestine"
/notes="Cloning vector pME18SFL3"
135..1289
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB15404.1"
/db_xref="GI:10439020"
/translation="MSENQTMLEPILHGGTSYGRGIRVLPESNDVDOLKKDLES
ANLTANOHSVDVICOSEPDSPSSGSXLEVEKGCGLDLEFVEHGLDIDFDGATKTPD
KOKAPOORIRIRAKIRSKYRPRRLOSXKIHCKKLLDQVLEKRTPLCATL
VOLQNTSLYDKSKINTTKNOKRKAIVPVKNNGTLPVSNKCLLIEGCTLSEI
KFNVSIVPKSCHDELIDLSAPFLIGCTTHHCKSSLSKSTIONLSLVKTSWII
SSVADLGIIVPLOYVFWVTPTLDGTVLEAVLMDSKFPOIVASEVELMDLDOKSVI
MIMDFPCPGKIDAYPMLECFIKSYNVNTQDIOCYUJEDTVAEDVI"
BASE COUNT      555  a      319  c      334  g      544  t
ORIGIN

Query Match      66.6%; Score 1268.4; Dh 9; Length 1752;
Best Local Similarity 99.5%; Prtd. No. 8.1e-294;
Matches 1283; Conservative 0; Mismatches 6; Indels 1; Gaps 1:

Oy 616 gatttaagtcacatccatcgcgtctacaataatctgacataagacatttgaatcgaataac 675
Db 1 GATTTAAGTCACATCCATCGCGGTACAAA-CTGACAAATAGACATTTTAACTGACGATGAC 59
Oy 676 catgtccatgtgcaaaatctcttgaaggtlgaagacattcttgaatctatagccatcat 735
Db 60 CATGTTCATGTGGCAAGATCTCTGAAAGCTTGCAAGCTTCTTACAAATCTATAGCCCTTCAT 119
Oy 736 acccaactctcaaaagaattcagaatcagaatcagtaagttcagaatcagatcctc 795
Db 120 ACCGACTTCATATCAATCAATTCAGAGATCAGCAATGTTTAAAGTTTGAATCTTCACTT 179
Oy 796 catgagagttaccagttacagtcagtcagtcagtcagtcagtcagtcagtcagtcag 855
Db 180 CATGAGAGTACCATTAAGGCTCGGGGAATCAGGCTCTGCCAGAAATTAACCTCGATGCTG 239
Oy 856 gatacaactgaaaagaagattgaatctgcaaatcttgaacagcaatcagaatcagaatct 915
Db 240 GATACAACTGAAAAGATTAAAGATTGCAAAATTGACAGCAATTCAGCAATTCAGATGTT 299
Oy 916 atctgtcaatcagaacacagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 975
Db 300 ATCTGTCACTCAGACATCTGACGACAGCTTTCAGAGCTCTGGATCAGTATCATTAATACGAG 359
Oy 976 gtagaagaatgtcaaacagctatctgtcaaatcttacaatcagaatcagaatcagaatct 1035
Db 360 GTAGAAGAATGTCAAAAGCTATCTGTACAAATCTTACAGATCATCTAGTATTGGAGAGG 419
Oy 1036 acaccacatgtgcatcttgaacacaaaagctcctcaacataacacacacacacacacacac 1095
Db 420 ACACCATATGTGCGATTGTAACAAAGAGCTCTCTAACAAATACGCAATCCAGAGAAAA 479
Oy 1096 ttaggtcatataaagccagaagaactatctcagtcgtctaaatcagtcagtcagtcagtcag 1155
Db 480 TTGAGGTATATAAGCCCAAGACAGCTATTTCAGTCTGTTAACTGATTCAGCTTCCCTAATGT 539
Oy 1156 catctgtgcaagaagaatctcaacatgagagcagtcagtcagtcagtcagtcagtcagtcagtcag 1215
Db 540 CATTTGCTGCAAGAAAGTTCCACATGAGGCGGATTTGGAATTAATTTTTCAGAGTGTGTGA 599
Oy 1216 actaaacacccagctgtcgaagttcaacaaatcacatctatataatcacaacacacacacacac 1275
Db 600 ACTAAACCCAGATGTCCAGCTACAAATACATCTTATATATATTCACAAATATCTGAGAC 659
Oy 1276 actaaacacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1335
Db 660 ACTAAACACACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 719
Oy 1336 cagcttcaaaatgaatgtctactcttgcataagaagaagtcacacacacacacacacacacacacac 1395

/clone="HS102573"
/clone_lib="HS1"
/tissue_type="human small intestine"
/notes="Cloning vector pME18SFL3"
135..1289
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAB15404.1"
/db_xref="GI:10439020"
/translation="MSENQTMLEPILHGGTSYGRGIRVLPESNDVDOLKKDLES
ANLTANOHSVDVICOSEPDSPSSGSXLEVEKGCGLDLEFVEHGLDIDFDGATKTPD
KOKAPOORIRIRAKIRSKYRPRRLOSXKIHCKKLLDQVLEKRTPLCATL
VOLQNTSLYDKSKINTTKNOKRKAIVPVKNNGTLPVSNKCLLIEGCTLSEI
KFNVSIVPKSCHDELIDLSAPFLIGCTTHHCKSSLSKSTIONLSLVKTSWII
SSVADLGIIVPLOYVFWVTPTLDGTVLEAVLMDSKFPOIVASEVELMDLDOKSVI
MIMDFPCPGKIDAYPMLECFIKSYNVNTQDIOCYUJEDTVAEDVI"
BASE COUNT      555  a      319  c      334  g      544  t
ORIGIN

Query Match      66.6%; Score 1268.4; Dh 9; Length 1752;
Best Local Similarity 99.5%; Prtd. No. 8.1e-294;
Matches 1283; Conservative 0; Mismatches 6; Indels 1; Gaps 1:

Oy 616 gatttaagtcacatccatcgcgtctacaataatctgacataagacatttgaatcgaataac 675
Db 1 GATTTAAGTCACATCCATCGCGGTACAAA-CTGACAAATAGACATTTTAACTGACGATGAC 59
Oy 676 catgtccatgtgcaaaatctcttgaaggtlgaagacattcttgaatctatagccatcat 735
Db 60 CATGTTCATGTGGCAAGATCTCTGAAAGCTTGCAAGCTTCTTACAAATCTATAGCCCTTCAT 119
Oy 736 acccaactctcaaaagaattcagaatcagaatcagtaagttcagaatcagatcctc 795
Db 120 ACCGACTTCATATCAATCAATTCAGAGATCAGCAATGTTTAAAGTTTGAATCTTCACTT 179
Oy 796 catgagagttaccagttacagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 855
Db 180 CATGAGAGTACCATTAAGGCTCGGGGAATCAGGCTCTGCCAGAAATTAACCTCGATGCTG 239
Oy 856 gatacaactgaaaagaagattgaatctgcaaatcttgaacagcaatcagaatcagaatct 915
Db 240 GATACAACTGAAAAGATTAAAGATTGCAAAATTGACAGCAATTCAGCAATTCAGATGTT 299
Oy 916 atctgtcaatcagaacacagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 975
Db 300 ATCTGTCACTCAGACATCTGACGACAGCTTTCAGAGCTCTGGATCAGTATCATTAATACGAG 359
Oy 976 gtagaagaatgtcaaacagctatctgtcaaatcttacaatcagaatcagaatcagaatct 1035
Db 360 GTAGAAGAATGTCAAAAGCTATCTGTACAAATCTTACAGATCATCTAGTATTGGAGAGG 419
Oy 1036 acaccacatgtgcatcttgaacacaaaagctcctcaacataacacacacacacacacacac 1095
Db 420 ACACCATATGTGCGATTGTAACAAAGAGCTCTCTAACAAATACGCAATCCAGAGAAAA 479
Oy 1096 ttaggtcatataaagccagaagaactatctcagtcgtctaaatcagtcagtcagtcagtcag 1155
Db 480 TTGAGGTATATAAGCCCAAGACAGCTATTTCAGTCTGTTAACTGATTCAGCTTCCCTAATGT 539
Oy 1156 catctgtgcaagaagaatctcaacatgagagcagtcagtcagtcagtcagtcagtcagtcagtcag 1215
Db 540 CATTTGCTGCAAGAAAGTTCCACATGAGGCGGATTTGGAATTAATTTTTCAGAGTGTGTGA 599
Oy 1216 actaaacacccagctgtcgaagttcaacaaatcacatctatataatcacaacacacacacacac 1275
Db 600 ACTAAACCCAGATGTCCAGCTACAAATACATCTTATATATATTCACAAATATCTGAGAC 659
Oy 1276 actaaacacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1335
Db 660 ACTAAACACACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 719
Oy 1336 cagcttcaaaatgaatgtctactcttgcataagaagaagtcacacacacacacacacacacacacac 1395

RESULTS
7
HSM800212
LOCUS      HSM800212      2012 bp      mRNA      PRI      18-FEB-2000
DEFINITION      Homo sapiens mRNA: CDNA DKFZP586D211 (from clone DKFZP586D211);
partial cds.
ACCESSION      AL050120
VERSION      AL050120.1 GI:4884145
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2012)
AUTHORS      Wambute,K., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE      Submitted (10-MAR-1999) M1PS, Am Klopferspitz 18a, D-82152
JOURNAL
COMMENT
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ). Email S. Wiemann@dkfz-heidelberg.de;
sequenced by ACOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZP586D211) is available at the RZPD in Berlin.
Please contact the RZPD: Resourcencentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
Location/Qualifiers
1..2012
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP586D211"
/clone_lib="586 (synonym: huet1). Vector pSport1; host
DH10B; slices Ncl1, Sma1/Mlu1"

```

```

/dev_stage="adult"
/tissue_type="uterus"
/map="530.1 CR from top of Chr7 linkage group"
1..411
/gene="DKF2p586D211"
<1..411
/gene="DKF2p586D211"
/note="WUGSC:H_DJ0907C10.1, partly"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB43281.1"
/db_xref="GI:4884146"
/translation="DHGCKKCCSSLRSLQNLNVKTSWIPSSVAEALGIVPLQYVF
WMFLDDGTGLAYLMDSDKFFQIPASEVLMDDLOKSDVMIMDFCPGGIKIDAY
PWLBCFKSYNVTNGTDNQICYQIFDTTVAEDVT"
polyA_signal 1965..1970
polyA_site 1983
BASE COUNT 631 a 343 c 357 g 681 t
ORIGIN

```

```

Query Match 21.4% Score 408.4; DB 9; Length 2012;
Best Local Similarity 99.8%; Pred. No. 1.5e-87;
Matches 409; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1495 atcaactatgatgtaaacagtggttagttgagatccatacaaaatctaattccctgg 1555
|||||
Db 2 ATCAGTATGGATGTAACAGTGTTCTAGTTTGAGATCCATACAAATCTAANTTCCCTGG 61

QY 1556 ttgataaaacatcgtgggttcctctctctctggtggcagaagcactgggtattatccctcc 1615
|||||
Db 62 TTGATAAAACATCGTGGATCTCTCTCTGTGGCAGAGCACTGGTATTCTACCCCTCC 121

QY 1616 aatatatatttattatgacatttaccattgtaataagaaacaggaatfactaaagacattatc 1675
|||||
Db 122 AATATCTGTGTGTTATGACCTTTACCTTGATGATGATGCAACAGAGTACTAGAGCTTATC 181

QY 1676 tcatgattctgacaaattcttcagattccagatccagaaagttctatagatgatatnacc 1735
|||||
Db 182 TCATGGATCTGCACAAATCTTCCAGATTCCAGATCCGAAATGTTCTGATGATGATGACC 241

QY 1736 ttcaaaaaatgtagatgataatgataatgattttttctctcccaagaaatdaaaattgatg 1795
|||||
Db 242 TTACAGAAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301

QY 1796 cataccctgattgagatgattctcaagtcatacaaatgtcacaaatgaaacagaaataatc 1855
|||||
Db 302 CATATCCGTGGTGGATGCTTCATCAAGTCATACAAATGTCACAAATGGAAACAGATAATC 361

QY 1856 aaattgctatcagatttttgcacaccacagttgacagaaagtgaattctaa 1905
|||||
Db 362 AATTGCTATCAGATTTTGCACACCACAGTTGCGAGAGATGTANTCTAA 411

```

```

RESULT 8
AC004925 AC004925 bp DNA PR1 21-DEC-1999
DEFINITION Homo sapiens PAC clone RP5-907C10 from 7431-3412, complete
sequence.
ACCESSION AC004925
VERSION AC004925.1 GI:4156174
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 154959)
AUTHORS Jones,K., Tin-Wollam,A. and Keppler,D.
TITLE The sequence of Homo sapiens PAC clone RP5-907C10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 154959)
AUTHORS Waterston,R.H.
TITLE Direct Submission

```

```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 154959)
Waterston,R.H.
Direct Submission
Submitted (14-JAN-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 154959)
Waterston,R.
Direct Submission
Submitted (22-JAN-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 154959)
Waterston,R.
Direct Submission
Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 14, 1999 this sequence version replaced gi:3213072.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_DJ0907C10

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NIGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/DTB/CHR7, send
mailto:cgreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://haicpac.mad.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong. VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP5-907C10; actual end is at 154959 of RP5-907C10.

The sequence from 4689 to 5920 is from PCR product from PAC RP5-907C10 DNA.

```

FEATURES
Location/Qualifiers
1..154959
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q31-3q32"

```

repeat_region	/clone-"RP5-907C10"
repeat_region	/clone_id-"RPCL-5"
repeat_region	66..3259
repeat_region	/rpl_family-"L1"
repeat_region	731..752
repeat_region	/rpl_family-"AT-rich"
repeat_region	1128..1429
repeat_region	/rpl_family-"Alu"
repeat_region	1652..1791
repeat_region	/rpl_family-"L1"
repeat_region	1810..1936
repeat_region	/rpl_family-"L1"
repeat_region	1949..2296
repeat_region	/rpl_family-"L1"
repeat_region	2297..2595
repeat_region	/rpl_family-"Alu"
repeat_region	2596..2760
repeat_region	/rpl_family-"L1"
repeat_region	<2816..25850
repeat_region	/gene-"WUGSC:H_DJ0907C10.1"
repeat_region	join<2816..2870.6849..7005.8634..8839.14398..14533.20470..20568.22507..22598.24455..24560.26738..25480>
repeat_region	/note-"WUGSC:H_DJ0907C10.1"
repeat_region	/note-"supported by human ESTs H23395 (NID:q892090) and A412636 (NID:q1687976), mouse ESTs W83982 (NID:q1394952) and A4717633 (NID:q2729907), and GenScan; H_DJ0907C10.1"
repeat_region	/cdon_start-1
repeat_region	/evidence-not_explicit
repeat_region	/proct_id-"AB06852.1"
repeat_region	/id_xref-"GI:4176376"
repeat_region	/transcript-"SCSVALVEVER7G61.SAF.11.THOYLDKPTTLCALIKOKALNOYR
repeat_region	LSK1LTKNKKRILPQSVHJHCPCRHLDLVEVHECDL11PQSGATKTHVAVLQMSLY
repeat_region	IAK1LTKNKKRKAQVAHVHKNNG11PLSNCLDL1EGGTLSLCKNSKFNVS1IPK
repeat_region	SHEDLELDLAPPL1GCT1HNYGCKSSLSR1JONLSLDRKSM1PSSVALD1C
repeat_region	VPLQYFVETFLDGGTCGLVLEATVLMDSKFPOTFASVEVLMDD1QKSDM1MDMFCIPP
repeat_region	G1K1ADYPM1KCF1KSYNVNCTN1G1CYQ1PHTTVADV1"
repeat_region	2844..2870
repeat_region	/gene-"WUGSC:H_DJ0907C10.1"
repeat_region	/note-"match to EST H23395 (NID:q892090) ym57q06.1"
repeat_region	3365..3397
repeat_region	/rpl_family-"AT-rich"
repeat_region	3436..3544
repeat_region	/rpl_family-"L1"
repeat_region	3569..3674
repeat_region	/rpl_family-"(CAT)u"
repeat_region	3684..4963
repeat_region	/rpl_family-"L1"
repeat_region	5057..5355
repeat_region	/rpl_family-"Alu"
repeat_region	5384..5685
repeat_region	/rpl_family-"Alu"
repeat_region	6265..6308
repeat_region	/rpl_family-"purine-rich"
repeat_region	6628..6748
repeat_region	/rpl_family-"L2"
repeat_region	6849..7005
repeat_region	/gene-"WUGSC:H_DJ0907C10.1"
repeat_region	/note-"match to EST H23395 (NID:q892090) ym57q06.1"
repeat_region	6939..7005
repeat_region	/gene-"WUGSC:H_DJ0907C10.1"
repeat_region	/note-"similar to Mus musculus EST W83982 (NID:q1394952) nt35f06.1"
repeat_region	7094..7121
repeat_region	/rpl_family-"AT-rich"
repeat_region	7421..7446
repeat_region	/rpl_family-"AT-rich"
repeat_region	7829..7862
repeat_region	/rpl_family-"AT-rich"
repeat_region	8174..8194
repeat_region	/rpl_family-"AT-rich"
repeat_region	8634..8839
repeat_region	/gene-"WUGSC:H_DJ0907C10.1"
repeat_region	/note-"similar to Mus musculus EST W83982 (NID:q1394952)

	misc_feature	mF35106..r1"	/gene-"MUGSC:H.DJ9407C10.1"	/note-"match to EST H23395 (NID:g892090) ym57g06..r1"
	misc_feature	718.._R839	/gene-"MUGSC:H.DJ9407C10.1"	/note-"similar to Mus musculus EST AA717633 (NID:g2729907)
	repeat_region	B935...B981 /rpt_family-"AF-rich"	bp94d10.r1"	
	repeat_region	9616...9810 /rpt_family-"L1"		
	repeat_region	9834...9891 /rpl_family-"L1"		
	repeat_region	9941...10115 /rpl_family-"L1"		
	repeat_region	10320...10728 /rpl_family-"L1"		
	repeat_region	10526...10907 /rpl_family-"L1"		
	repeat_region	10908...11201 /rpl_family-"Alu"		
	repeat_region	11202...12306 /rpl_family-"L1"		
	repeat_region	12517...12741 /rpl_family-"MIR"		
	repeat_region	12854...12985 /rpl_family-"I2"		
	repeat_region	13001...13156 /rpl_family-"MIR"		
	repeat_region	13466...13481 /rpl_family-"L1"		
	misc_feature	14396...14533 /gene-"MUSGC:H.DJ10907C10.1" /note-"Similar to Mus musculus EST AA717633 (NID:g2729907) vp94d10.r1"		
	misc_feature	14396...14533		
Coverage Match	11.0%;	Score 208.6; Fr 9;	Dlength 15495%;	
Hest Local Similarity	93.9%;	Pred. No.	J.5e-39;	
Matches 217:	Conservative 0;	Mismatches 14;	Indels 0;	Gaps 0;
OY	1140 tcatatccctaaagattcatalgtacataagaaggcttcgcatagaagcaagttagatatcat 1159			
Db	8610 TCATTTCATTTCCCAATAAATTTCAGCGACGAACTTCCACAATCACGGCATTTTGCAATTAAT 8669			
OY	1200 ttttggadtaqatqaactaaaaccrcratlqtcaaqlaacnaalacalcataltaa 1279			
Db	8670 TTTTTGAGCATGTGCCAACAAAACCAGATCTCAACTACAAAAATTCATCATTAATGA 8729			
OY	1260 ttcgaaaaatcggcccctcabaanaatlcaaaaanraqdaaatlaacaqtlccaalltgtgae 1319			
Db	8730 TTCAAAATCTCGACACCACTAAMAAATCAAAGAAGAAAGTACCACTCATTTTGGA 8789			
OY	1320 aaalatgatattctccrcqatltleaatqaabqlctacltttgatadaaq 1370			
Db	8790 AATTAATGCTATTCTCCCCTTTGCMAATGCAATTCCTACTTTTGATAGAGG 8840			
RESULT	6			
Locus	G12915	320 bp	RNA	STS
DEFINITION	SMSZS290 Eric D. Green Homo sapiens STS genomic sequence tagged clone			28-Sep-1998
ACCESSION	G12915			
VERSION	G12915.1			
KEYWORDS	STS,			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. Houliard G.G., Jyot M., Idejima K., Braden V.V., Cunningham A.P., AUTHORS			

[illegible]

```

gene-      "PRLSR3.MLS"  
1741..2862  
/gene-"MYPU_0020"  
1741..2862  
/gene-"MYPU_0020"  
/standard_name-"dun"  
/evidence-not_expimental  
/note-"Identified by sequence similarity; putative;  
similar to SWISSPROT:P34029 (DP3H_SPLIC) blastp2 p=3e-14  
C=4.8%"  
/codon_start-1  
/transl_table-4  
/evidence-not_expimental  
/product-"DNA POLYMERASE III, BETA CHAIN"  
/protein_id-"CAC13176.1"  
/db_xref-"GI:14089415"  
/translation-"MELFLKRSHEPLLELPNPKLLESNN:NYELGCAVPOVEHQDILII  
SSNDISILNLJHERFDIKNKEPCESDPIVLNSLKNI LKGSNSILREKPNLLK  
VNLSENNLTITLVETNRKLAQOK IETPSOENPTLNKYVRELISLSMSSTILLST  
GTPEKL KSNGLTLTKTVLEIIVMENVPFKNFVLIIDRKLSSLIDKYSIVNDIK  
KRNKRI LEYNISRKERKRLKLSYWDTADFSVEFSDNEVSESLIKPLIMANYLRK  
IINVDKMLSIPTTENDIRVVISSEPMNKROIALALRGH"  
2864..3085  
/gene-"MYPU_0040"  
2864..3085  
/gene-"MYPU_0040"  
/note-"Identified by sequence similarity; putative;  
similar to SWISS-PROT:P05050 (YAAU_BAC50) blastp2 p=6e-06  
C=4.5%"  
/codon_start-1  
/transl_table-4  
/evidence-not_expimental  
/product-"conserved hypothetical protein"  
/protein_id-"CAC13176.1"  
/db_xref-"GI:14089416"  
/translation-"MTILKGEPLTGOLIRKRLIDSGOGIKPFMOASHOVINKCKL  
NTKNIKVKOCRTWIDPOLYTIKAONNE"  
3088..3771  
/gene-"MYPU_0040"  
3088..3771  
/gene-"MYPU_0040"  
/note-"Identified by sequence similarity; possibler  
similar to SWISS-PROT:P43035 (YL11_MYCHO) blastp2 p=1e-28  
C=5.3%"  
/codon_start-1  
/transl_table-4  
/evidence-not_expimental  
/product-"conserved hypothetical protein"  
/protein_id-"CAC13177.1"  
/db_xref-"GI:14089417"  
/translation-"MKYIAKSKLHKRVFRKLSHVLFYKGFINRYLAGFYNGIKVOILRI  
KINDLVNIEDEKKYALKNNEDILPVNKFVVIKKMFGEHFKNRKSLLKNTYDLKKFWK  
QINDLAKNMWLIVETKRDIIPRYTSLVERKIDPFNDVIGHDLNFKNILANLNKNEVLIIDPE  
WKKGSKSYVDICITYKFNINIKNELIKRDLINKEFTDYILVMYKFIIEAELEKYSKSK  
SISPWTEK"  
3774..4422  
/gene-"MYPU_0050"  
3774..4422  
/gene-"MYPU_0050"  
/note-"no similarity found"  
/codon_start-1  
/transl_table-4  
/evidence-not_expimental  
/product-"unknown; predicted coding region"  
/protein_id-"CAC13176.1"  
/db_xref-"GI:14089418"  
/translation-"MEHWVFSEFOIGECNNLPYSLAFETSLSNKGRILDKKHIFGV  
VMSLDISKNLKKRYNNVLDNSNSLNYNFLKPRVLDKSNMINVSFVNINILIPST  
LAKISPSLAKULIYUINKRKELKNSSLPKRIPLNLIYINAIYISYESKHGGP  
VLSNSENTERKMWLINITFKTELIOFLMLKTRVILEEISHPKRYKCOYNAKIIP"  
4403..6750  
/gene-"MYPU_0060"  


```



```

TKMKATANIYLFNLADALATSTLPQSVNYLMCTWPCSTLICKLVISIDVYNNFTS
1FLLCTMSVDRIYAVCHPVRKALDRTPRNKR11NVCNMLSSA1GLVWFMWNTYXKQ
GSDCTTFTSHPTWYNELIKICVPIFAPIWPLIITVYCGMLILIKSKVMLSGSKR
KIDNLRRTTHWVAVVAVPIVCWEP1I1IYV1IKALVY1PETTFQVSMIFCLALGYN
SCILNPVYAPLDNFRKCFEPCFPTSSN1EGQNSNTR1KQNTBHHNSTANTYDKRHC
VKS1.
repeat_region 13353..13654 /note="AluIb repeat: matches 3..305 of consensus"
repeat_region 14114..14194 /note="L2 repeat: matches 2209..2295 of consensus"
repeat_region 14245..14354 /note="L1MAH repeat: matches 6206..6314 of consensus"
misc_feature 14726..15056 /note="match: GSS: Em:AO269347"
repeat_region 14355..14550 /note="MER58A repeat: matches 1..224 of consensus"
repeat_region 14615..15008 /note="L1MB1 repeat: matches 5633..6020 of consensus"
misc_feature 15644..16491 /note="match: GSS: Em:AO748231"
repeat_region 16425..16537 /note="L1MAN repeat: matches 6176..6289 of consensus"
repeat_region 16824..17072 /note="Alus9 repeat: matches 1..300 of consensus"
misc_feature 17700..18193 /gene="OPRM1"
/Note="match: GSS: Em:AO777968"
complement(17761..18197)
/gene="OPRM1"
repeat_region 18009..18044 /note="match: GSS: Em:AO805249"
misc_feature 18162..18373 /note="9 copies 4 mer 1919 97% conserved"
/Note="match: STS: Em:130054"
repeat_region 18505..18548 /note="11 copies 4 mer 9191 77% conserved"
repeat_region 18992..19300 /note="AluY repeat: matches 1..407 of consensus"
repeat_region 20485..20734 /note="L2 repeat: matches 2655..2705 of consensus"
repeat_region 21426..21734 /note="AluY repeat: matches 1..311 of consensus"
complement(22272..22843)
/gene="OPRM1"
/Note="match: GSS: Em:AO533470"
complement(22355..22858)
/gene="OPRM1"
/Note="match: GSS: Em:AO424676"
repeat_region 22427..22547 /note="MER69 repeat: matches 489..614 of consensus"
repeat_region 22575..22788 /note="MER69 repeat: matches 662..893 of consensus"
repeat_region 22769..23087 /note="MER69 repeat: matches 1991..2295 of consensus"
repeat_region 23088..23246 /note="FRAM repeat: matches 2..159 of consensus"
repeat_region 23247..23438 /note="MER69 repeat: matches 2295..2490 of consensus"
misc_feature 23763..24232 /note="match: GSS: Em:AO793055"
repeat_region 24067..24372 /note="AluIb repeat: matches 1..299 of consensus"
repeat_region 24601..25199 /note="L2 repeat: matches 1142..1818 of consensus"
repeat_region 25225..25512

```

Query Match 2.2% Score 42; Dh 9; Length 96310;

Best local Similarity 49.5% Pred No. 13; Mismatches 110; Indels 0; Gaps 0;

Matches 108; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

```

QY 1113 cagaagactatlcagcgtcgttaaacctccatcgcctcctaaatgctcgcagagagc 1112
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 22648 AGGAATCCCTGATMAAT7GTATTTGACAGCATTTGCCATGAAAACTTTATAGTCATGATA 22589
QY 1173 cccagatgaagcgcgtt11gaatataattttt1cagatggtcgaactaaaccgcagcgtcgt 1232
111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 22588 ATTGCATGTTACATGCTATATATCAGCTATATAGTCAGACGACGAAATGATMAAAATCC 22529
QY 1233 cgaattacgaat1ccatcattatataat11cgaatcgtcgaactaaaccgcagcgtcgt 1270
111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 22528 GTACTTACAAAAAATCTGCTATGATTAATMAAAATTT 22491

```

Search completed: Apr 11 9, 2002, 23:28:39
Job time: 1075.3 sec

Qy 1285 caaaaagqacqaaaaatagcaqcttcattttttgaaaaataaatggtattctctccqctttca 1344

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA-708
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1045 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE: 31J2
INDIVIDUAL ISOLATE: 31J2
US-08-960-780-20

Query Match 1.7%; Score 33.2; DB 4; Length 1045;
Best Local Similarity 49.4%; Pred. No. 9;
Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 1205 caaaagacgaaagtagcagtcatttttgcgaaataatagtgattctccgctttca 1344
DB 245 caaaaagcccaaaagaaacacagttgttcattttagaaaaagataaattagttcccatcaaa 304
QY 1345 atggaatgctacttttataagaaaggtacacactcgtatgaattttgcacactctcgaac 1404
DB 305 attgaatatcatctgcataaagcgttaaacgcacatagttcaaatgttttaagaatttgaaa 364
QY 1405 aagtttaatagtgatctctgtagatctgacacagcaagaaacctcgaacttttg 1458
DB 365 ttattttaaataatagttcaaaaacaaatctcagcaagtcgcaaacagacaggaatttg 418

RESULT 15

US-09-073-898-20
Sequence 20, Application US/09071898

Patent No. 6242669

GENERAL INFORMATION:

APPLICANT: Feltelson, Jerald S.

APPLICANT: Schnepl, H. Ernest

APPLICANT: Narva, Kenneth E.

APPLICANT: Stockhoff, Brian A.

APPLICANT: Schmeltz, James

APPLICANT: Lowler, David

APPLICANT: Dullum, Charles Joseph

APPLICANT: Muller-Gohn, Judy

APPLICANT: Stamp, Lisa

APPLICANT: Morrill, George

APPLICANT: Flusbad-Lee, Stacy

TITLE OF INVENTION: No. 6242669-9-1 Pesticidal Toxins and Nucleotide

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESS: Saliwanchik, Lloyd & Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL

COUNTRY: US

ZIP: 32606-6669

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/073,898

FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/960,780
FILING DATE: 30-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sanders, Jay M.
REGISTRATION NUMBER: 39,355
REFERENCE/DOCKET NUMBER: MA-708C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1045 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE: 31J2
INDIVIDUAL ISOLATE: 31J2
US-09-073-898-20

Query Match 1.7%; Score 33.2; DB 4; Length 1045;

Best Local Similarity 49.4%; Pred. No. 9;

Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 1285 caaaaagacgaaagtagcagtcatttttgcgaaataatagtgattctccgctttca 1344

DB 245 caaaaagcccaaaagaaacacagttgttcattttagaaaaagataaattagttcccatcaaa 304

QY 1345 atggaatgctacttttataagaaaggtacacactcgtatgaattttgcacactctcgaac 1404

DB 305 attgaatatcatctgcataaagcgttaaacgcacatagttcaaatgttttaagaatttgaaa 364

QY 1405 aagtttaatagtgatctctgtagatctgacacagcaagaaacctcgaacttttg 1458

DB 365 ttattttaaataatagttcaaaaacaaatctcagcaagtcgcaaacagacaggaatttg 418

Search completed: April 9, 2002, 22:21:04
Job time: 7638 sec


```

Email : filanq@llftech.com URL :
http://fulllength.lnvtirogen.com"
BASE COUNT      290 a    176 c    166 g    256 t    1 others
ORIGIN
Query Match      45.4% Score 883.8; DB 10; Length 889;
Best Local Similarity 99.6%; Pred. NO. 1 Mar-202;
Matches 885; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 871 qatttagaatctgcgaatattgacagcgcatafcacatcatctctcttcactcgtcaaa 930
|||||
Db 1 GATTTAGAAATCTGCAAAATTTGCACAGCCAAATCAGCATTCATCGATTCTCAATCAGAA 60
|||||
QY 931 ctctgacgcacagtcttccaaagctctggatcagctatcattatcacgaatctaaaaatctcaa 990
|||||
Db 61 CCTGACGACAGCTTTCCAAAGCTCTGGATCACTATCATATTACGAGGTAGAAGATCTCAA 120
|||||
QY 991 caactctctctacaactctacagatcatcaatctattctgaaaggacaccactctatgcc 1050
|||||

```

[illegible]

```

QY 1442 aaagccctggaaccllllaagacclllccaaaccccllcllcllaacgaagaaacatacact 1501
|||||
DB 421 AAGACCTGGAACTTTTGGACCTTTGACACACCAATTTCTTATVACAAGCAAAATACATCACT 480
QY 1502 atgagctgaacacagatgctcagctllagagatccaaagaaatclaaatllccctgllgala 1561
|||||
DB 481 ATGAGATGTAACAGCTTTTCACTTCAATCCATCAACAAATCTAAATTTCCCTGGCTCATTA 540
QY 1562 aaacacacgaagaccllcccllcllcllcllcllcllcllcllcllcllcllcllcllcll 1621
|||||
DB 541 AAACATGCTGGCAATTTCTCTCTCTCTGACAAACCACTCCGATATTTCTACCCCTCCAAATATG 600
QY 1622 tglcllcllcllcllcllcllcllcllcllcllcllcllcllcllcllcllcllcllcll 1681
|||||
DB 601 TCTTAATTTTATCACCTTTTACACCTTTGATCATGACAGATGATVACAAACCTTATCTCATGCG 660
QY 1682 atcttgaacaaatcttccaaatctccaaatccaaatcttgaatgaatgaatgaatgaat 1740
|||||
DB 661 ATTCTGCAAAATTTCTCTGACATTTCTGACATTCACGATTCCTGATGACTGATGACTTTGAC 720
QY 1741 aaagatgagaacllcllcllcllcllcllcllcllcllcllcllcllcllcllcllcllcll 1800
|||||
DB 721 AAACATGCTGGCAATTTCTCTCTCTGACAAACCACTCCGATATTTCTACCCCTCCAAATATG 780
QY 1801 cgcgt-gatllgaaatgcttccatgaaatccaaatccaaatccaaatccaaatccaaatccaa 1857
|||||
DB 781 CCGTGGCTTGGCAATTTCTGATCATGATCATGATCATGATCATGATCATGATCATGATCATGAT 838

RESULT 3
LOCUS BE891252 926 bp mRNA EST 20-OCT-2000
DEFINITION 601432077F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3917299 5',
mRNA sequence.
ACCESSION BE891252
VERSION BE891252.1 GI:10450799
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 926)
TITLE NIH-MGC Initiative: Primates; Catarrhini; Homiidae; Homo.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaups-f@mail.nih.gov
Tissue Procurement: ATCC/ICM/DP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The J.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Inyece Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the J.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LHAM9743 row: 3 column: 20
High quality sequence stop: 636.
location/Qualifiers
1. 926
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3917299"
/clone_1="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="p310b (phage-resistant)"
/notes="Organ: skin; Vector: pCMV-SORTO; Site_1: NotJ;
Site_2: SalI; Cloned unidirectionally. Primer: OJuo d1.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 273 a 205 c 212 g 238 t
ORIGIN
Query Match 32.1%; Score 611.2; DB 11; Length 926;

```

```

Hest local similarity 94.0%; Pred. No. 8,1e-137;
Matches 679; Conservative 0; Mismatches 38; Indels 5; Gaps 4;
QY 87 gaagctcltlaaaccccatatcctlaagaaagaaatgaatcttgcagttgtaact 146
|||||
DB 1 GAAGTCTTTAAGCCCCCATATCTAAGCAAGAAAGAACTGATTTTCTCTCACTGTAACAT 60
QY 147 tglgagcaagacaaatlaaaacclaaatllgcclgclcllcllcllcllcllcllcllcll 206
|||||
DB 61 TCTGCAACACCAAAATCTTAAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 120
QY 207 tccaaatgaatcttctaaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 266
|||||
DB 121 TCCAAATTAATTTATTAATAATTCAGATATTTCTTCTCTTCAACACTTGAATATTCAACATTA 180
QY 267 taagaaagaaagaaatcaagaaatcaagaaatcaagaaatcaagaaatcaagaaatcaagaaat 326
|||||
DB 181 TAAAGAAAGAAATTTCAACAGCTTATVACAGCTTCTGAGCTTCTGAGCTTCTGAGCTTCTGAG 240
QY 327 tttgaaagaaatcttctaaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 386
|||||
DB 241 TTTTCTGACACCCCTTATVACATVACATVACATVACATVACATVACATVACATVACATVACAT 400
QY 387 cctcgaatgaatgaagaaatcttctaaatgaatgaatgaatgaatgaatgaatgaatgaatga 446
|||||
DB 401 CCAAGAAATATVACAAAGCTTAACTGCTTTGAGCAATCTGATGATGATGATGATGATGATGAT 460
QY 447 atcttgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 506
|||||
DB 361 ATTAATTAATAATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 507 gaaagaaagaaagaaatgaagaaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 566
|||||
DB 421 GCGCAAGCAAGATCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 480
QY 567 atcgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 626
|||||
DB 481 ACCATTTTCAATTTCTGAGATCTTAAATACAAAGCTTTCTTGAAGCTTGAAGCTTGAAGCT 540
QY 627 caatcgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 686
|||||
DB 541 CACCAATCTGATCTGAGATCTTAAATACAAAGCTTTCTTGAAGCTTGAAGCTTGAAGCTTGA 596
QY 687 g-cgaagaaatcttctaaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 745
|||||
DB 600 GCGCCGATCTCTGAGATCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 805
|||||
DB 657 CACATGCAATTTCTGAGATCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 716
QY 806 cc 807
DB 717 CC 718

RESULT 4
LOCUS B1257710 1159 bp mRNA EST 17-JUL-2001
DEFINITION 602967368F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5106881 5',
mRNA sequence.
ACCESSION B1257710
VERSION B1257710.1 GI:14813351
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1159)
TITLE NIH-MGC Initiative: Primates; Catarrhini; Homiidae; Homo.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

```

Query Match	30.8%	Score 587;	DB 11;	Length 1159;
Host Local Similarity	91.1%	Pred. No. 5,60-131;		
Matches 706;	Conservative	0;	Mismatches 50;	Indels 19; Gaps 7;
766	capacuatgttaadttlaagaatttctatctcatatgaattaccatctacgttcgaggaatc	825		
2	CAGACAATGTTAACTTTTAGAGTTTCATCTCATGGAGGTACCAGTTTAGCTCGCGGAATC	61		
826	jqggtcttgccagaaataactctgattgaatcaactgaataaataatttamaatctaca	885		
62	AGGTTCTTCGACGAAGAATCACTCTGATGTGGATCAACTGAAAAGGATTTAGAAATC	121		
886	aattgcagacgaatacagcatctcaagatgttatctctcaatcagaacctcagacagcttt	945		
122	AAATTTGACAGGCAATCAGCATTCAGATGTTATCTGTCAATCAACAGCTGACGACACGCTTT	181		
946	ccaagctctggatcagtatcatctatccagagtgaaagagtgcaacagctatctctaca	1005		
182	CCAAGCTCTGGATCAGTATCATTTATACGAGGTAGAAAGATGTCAACAGCTATCTGCTACA	241		
1006	atactctcaagatcatcagctatttgagagagagacacactatagcattctctgaacaaata	1065		
242	ATACTTTACAGATCATCAGTATTTGGAGAGACACACACTATGTGCCATTTTGAACAAAAA	301		
1066	gctctctcaagatatccgcatactcagagcaaatctgaattctataaagccagagagctctct	1125		
302	GCTCTCAACAAATACCGCATCCGAGCAAAATTCAGATTCATATTAGCGCCAGAGCTATTTT	361		
1126	caatctgttaaaactctatggctataatgca-ttctctcagaagaagttccacataaaggg	1184		
362	CAGTCTGTGTTAAACTTCATTTGCCCTAAAGTGTATTTTGTCTGCAAGAAAGTTCCACATCAGGG	421		
1185	cga-ttctgatatataattctctgaatggtacactataaccccaattctcagttcacaa	1243		
422	CGATTTTGGATTAATTTTTCAGATGCTGTCNA-TTAAAGCCGAGATGTCAAGCTACAAA	481		
1244	atacatcatctatatgatctcaaaaatctcgaccactataaatac-aaataagacaaagaata	1302		
482	ATACATCATTTATATGATTTCAAAAATCTGGAGCACATAAAAATTCGACAAAGCCGCAAAAGTA	541		
1303	gcagttcatttttg-aaaaataatggatctctccgcctttcgaatgaatgctactct	1360		
542	CGAGTTCAATCTGTGCACACAATAATGGTTATCTCCCGCTTTTCAATGAATGCTACTTTT	601		
1361	tgaagaaggaagtacactcaatgaaa-tttgcaactctctcgaaacagtttaataagtata	1419		
602	TCAAGAAGGAGGTACACTCAGTGAATTTTGGAACTCTCGAAACAGTCTAATATGGTA	661		

[illegible]

```

1. .588
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3912010"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notice="Organ: uterus; Vector: pCMV SPORT6; Site_1: NOF1; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.1 kb."
207 3 115 c 105 g 160 t

```

BASE COUNT
ORIGIN

Query Match	30.6%	Score 542.8	DB 11	Length 588
Best. Local Similarity	99.7%	Prod. No. 5,44-130		
		0 mismatches		
		0 indels		0 gaps

[illegible]

(HXSE-) HXSEO INC.

PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D:
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J:
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT:

XX WPI: 2001-442253/47.

DR N-PSDB: AA159281.

XX Novel nucleic acids and polypeptides, useful for treating disorders
PI such as central nervous system injuries -

XX Example 5: SEQ ID NO 3270: 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA15799) and
CC the encoded polypeptides (AA158542-AA158543) with neurotrophic
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy, and
CC localized neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager syndrome. Other uses include the
CC utilization of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 524 AA:

Query Match 96.5%; Score 2637; DB 22; Length 634;
Best Local Similarity 99.2%; Pred. No. 9, 5e-263;
Matches 503; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

0Y 1 MSLVFAINVIYPLNOLKGGTIIVNWGVKFFKPPYLSKGTIDYCSVTVIVDQTNVKTCL 50
DB 1 MSLVFAINVIYPLNOLKGGTIIVNWGVKFFKPPYLSKGTIDYCSVTVIVDQTNVKTCL 50
0Y 61 LFSGNVEALPIYKNGDIYFRHLKIOVKKETOGITSSGFASLFEGLTGAPIIPRIS 120
DB 61 LFSGNVEALPIYKNGDIYFRHLKIOVKKETOGITSSGFASLFEGLTGAPIIPRIS 120
0Y 121 KTFNFTEDHKVVEALRWASTHMSKWLKLDVQPMOYEDLTQILGKAEVQDASFL 180
DB 121 KTFNFTEDHKVVEALRWASTHMSKWLKLDVQPMOYEDLTQILGKAEVQDASFL 180
0Y 121 KYFNFTEDHKVVEALRWASTHMSKWLKLDVQPMOYEDLTQILGKAEVQDASFL 180
DB 121 KYFNFTEDHKVVEALRWASTHMSKWLKLDVQPMOYEDLTQILGKAEVQDASFL 180
0Y 181 LKWDGTFPFPSPWVYLIDVLEGDLSHHPLONLIDILVYDHNHVAARSILKVGSLP 240
DB 181 LKWDGTFPFPSPWVYLIDVLEGDLSHHPLONLIDILVYDHNHVAARSILKVGSLP 240
0Y 241 IYSLHTKIOSMNSENDTMSLEFHLHGCTISYGCIRVLPESNDVQOLKDLFSAULTAH 300
DB 241 IYSLHTKIOSMNSENDTMSLEFHLHGCTISYGCIRVLPESNDVQOLKDLFSAULTAH 300
0Y 301 OHGCDVCGSEPPDSFSSGVSLSYEVRGCOOLATILIDHOYLERPLCAILKOKAPQOY 360
DB 301 OHGCDVCGSEPPDSFSSGVSLSYEVRGCOOLATILIDHOYLERPLCAILKOKAPQOY 360
0Y 361 RIKAKLRSYKPRRLFOGVKLHCPKCHLLOEVRPHEGDLIDFDGATKTPDVKQNTSLYD 420
DB 361 RIKAKLRSYKPRRLFOGVKLHCPKCHLLOEVRPHEGDLIDFDGATKTPDVKQNTSLYD 420
0Y 421 SKITWTKNGKGRVAAHVEVNNKGLPLSRCLLIEGGTISEICKLSNRKRSVFEVSGH 480
DB 421 SKITWTKNGKGRVAAHVEVNNKGLPLSRCLLIEGGTISEICKLSNRKRSVFEVSGH 480
0Y 481 EDLELDLSAPFLHOGTIIHYGTGYCT 507
DB 481 EDLELDLSAPFLHOGTIIHYGTGYCT 507

RESULT 2

AA893478 ID AA893478 standard; Protein: 634 AA.

XX AA893478:

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12761.

XX Human: primer: detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000: 2000EP-0116126.

XX 29-JUL-1999: 99JP-0248036.

XX 27-AUG-1999: 99JP-0300253.

XX 11-JAN-2000: 2000JP-0118776.

XX 02-MAY-2000: 2000JP-0183767.

XX 09-JUN-2000: 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J:

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T:

XX WPI: 2001-318749/34.

XX Claim 8: SEQ ID 12761: 2537P + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the combination of
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences: AA892446 to
CC AA895893 represent human amino acid sequences: and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 534 AA:

Query Match 96.5%; Score 2637; DB 22; Length 634;
Best Local Similarity 99.2%; Pred. No. 9, 5e-263;
Matches 503; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

0Y 1 MSLVFAINVIYPLNOLKGGTIIVNWGVKFFKPPYLSKGTIDYCSVTVIVDQTNVKTCL 60

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:08:13 ; Search time 58.75 Seconds
(without alignments)
653.106 Million cell updates/sec

Title: US-09-816-248-17
Perfect score: 2732
Sequence: 1 MSIVPATNRYITPLNOLKGG.....HHYGTGCTPICVCYDLYT 518

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2637	96.5	634	22	Human Polypeptide
2	2637	96.5	634	22	Human protein sequ
3	2637	96.5	634	22	Human protein sequ
4	2628	95.2	634	22	Human protein sequ
5	922	33.7	307	22	Human polypeptide
6	155	5.7	30	22	peptide -891 enco
7	125	4.6	467	21	Arabidopsis thalia
8	105	3.9	2391	15	Cardamom-1-phosphat
9	105	3.8	555	17	Serine kinase SRK
10	104	3.8	1112	21	Haemo: nagic enter
11	102	3.7	4393	22	Amino acid sequenc
					AA1980
					AA1981
					AA1982
					AA1983
					AA1984
					AA1985
					AA1986
					AA1987
					AA1988
					AA1989
					AA1990
					AA1991
					AA1992
					AA1993
					AA1994
					AA1995
					AA1996
					AA1997
					AA1998
					AA1999
					AA2000
					AA2001

12	99	3.5	684	20	AAV34460
13	99	3.6	686	20	AAV34336
14	99	3.6	1002	21	AA38741
15	98.5	3.6	653	22	AA47250
16	98	3.5	872	22	AAU07149
17	98	3.6	372	19	AAW98360
18	97	3.5	372	22	AAW60318
19	97	3.6	1123	21	AA35865
20	96	3.5	1139	21	AA18793
21	95.5	3.5	332	21	AA59490
22	95.5	3.5	354	21	AA59488
23	95.5	3.5	357	21	AA59488
24	94.5	3.5	1568	20	AAV13462
25	94.5	3.5	1568	21	AAW28522
26	94.5	3.5	1568	22	AAU00239
27	94.5	3.5	1568	22	AAU00435
28	94.5	3.5	1568	22	AAW70131
29	94.5	3.5	1568	22	AAW1693
30	94	3.4	467	21	AAW25071
31	94	3.4	536	21	AAW18793
32	94	3.4	537	21	AAW23070
33	94	3.4	557	21	AAW31369
34	94	3.4	606	21	AAW31372
35	94	3.4	626	21	AAW31371
36	94	3.4	1139	21	AAW18794
37	94	3.4	2285	20	AAW98149
38	93.5	3.4	519	21	AAW41694
39	93.5	3.4	1461	19	AAW64468
40	93.5	3.4	1461	22	AAW90743
41	93	3.4	1119	20	AAW08114
42	93	3.4	1119	21	AAV70674
43	93	3.4	1119	22	AAV12347
44	93	3.4	1119	22	AAU00827
45	93	3.4	1119	22	AAW80265

ALIGNMENTS

RESULT 1	AAW40125	standard; Protein: 634 AA.
ID	AAW40125	
XX	AAW40125	
AC	AAW40125	
DI	22-OCT-2001	(first entry)
DE	XX	Human polypeptide SEQ ID NO 3270.
XX	XX	
XX	XX	Human: neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
KW	KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW	KW	leukaemia.
CS	XX	Homo sapiens.
FN	XX	W0200153312-A1.
XX	XX	
PD	XX	26-JUL-2001.
XX	XX	
PF	XX	26-DEC-2000: 2000WD-US34263.
XX	XX	
PR	XX	21-JAN-2000: 2000US-0488725.
PR	XX	25-APR-2000: 2000US-0552317.
PR	XX	09-JUL-2000: 2000US-0598042.
PR	XX	19-JUL-2000: 2000US-0620312.
PR	XX	03-AUG-2000: 2000US-0653450.
PR	XX	14-SEP-2000: 2000US-0662191.
PR	XX	15-OCT-2000: 2000US-0693036.
PR	XX	25-NOV-2000: 2000US-0727344.
XX	XX	

Porphyromonas ging
Arabidopsis thalia
Human p7. Homo s
C. elegans CRIM F
H. pylori GHP. 134
Helicobacter fytol
Arabidopsis thalia
Amino acid sequenc
Arabidopsis thalia
Arabidopsis thalia
Viral-encoded sema
Human VESPR. Homo
Viral-encoded sema
Human VESPR. Homo
Amino acid sequenc
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Amino acid sequenc
Bacillus sub.115
Human ORFX OF1459
Human secreted pro
Human CM420.2 pro
Human PRO326 prote
Human PRO326 polyp
Human immune respo
Human PRO325 prote

us-09-816-248-17.rsp

[illegible]

```

RESULT 14
ID 09YV4A PRELIMINARY:
09YV4A: Created)
AC 09YV4A-1999 (IREMBLrel. 12, last sequence update)
DI 01-NOV-1999 (IREMBLrel. 17, last annotation update)
DT 01-JUN-2001 (IREMBLrel. 17, last annotation update)
DE KIAA0580 PROTEIN (FRAGMENT).
GN KIAA0580.
NM Homo sapiens.
OS Chordata: Vertebrata: Euteleostomi:
OC Eukaryota: Metazoa: Primates: Catarrhini: Homiidae: Homo.
OC Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
OX NCBI_TaxID=9605.
RN (1)
SEQUENCE FROM N.A.
RP TISSUE=BRAN: pubMed=5628581; Tanaka A., Kotani H., Nomura N.
RC MEDLINE=98290545; Miyajima N.,
RX Nagase T., Ishikawa K.,
PA Ohara O.,
FA "Prediction of the coding sequences of 100 new cDNA clones from human which can
RI "Prediction of the coding sequences of 100 new cDNA clones from human which can
RL code for large proteins in vitro."
RT DNA Res 5:31-39(1998).
EMBL: AB01152; BAA25506.1;
DR EMBL: AF000198; ZNF-GCS.
DR InterPro: IPR001164;
DR InterPro: IPR001849; PH.
DR InterPro: IPR00159; PH. 4.
DR Pfam: PF00159; RHOGAP: 1.
DR Pfam: PF00620; RHOGAP: 1.
DR Pfam: PF01412; RHOGAP: 1.
DR Pfam: PF00405; REVINTACTING.
DR PRINTS: PSS0003; PH_DOMAIN: 3.
DR PROSITE: PSS0003; RHOGAP: 1.
DR SMART: SM00233; PH: 3.
DR SMART: SM00324; RHOGAP: 1.
DR SMART: SM00324; RHOGAP: 1.
FT NON_TER
SO SEQUENCE 1044 AA: 119463 MW: 58822456DEMAA53BD CPO64:
Query Match 4.0% score 110.5; DB 4: length 1044:
Best Local Similarity 20.2% Pred. No. 1.7:

```

Query Match	4.0%;	Score 10
Best Local	30.8%;	Score 10
Matches	110	

	Conservative	Pred. No. 3.5;	DB 5;	Length 1217;
Oy	45	SVTVIVDOTVNYKLTCL	86;	Mismatches 166; Indels 172; Gaps 30;
Db	483	AVYSLVDLOVNAHNSIIIVKRLIRKRYKGIYVNSVETLISTAKIAS		
Oy	52	ETOGIISGSFASLTGEGIGAPIRIS		
Db	539	TAERITYS		
Oy	144	MSPEVILLKCDVOPMOYEDTLTOLLGKAEVDSGASFLKVNDDG		
Db	589	MFPDGETISIM		
Oy	151	FPSKREVLJODL		
Db	641	DALIKRIGISTAYLE		
Oy	229	VASLSLVGS		
Db	696	LEELLEAFADOLNGFASLPNOSLTIEIMARVEDEAKIKIGIGDPRKSAWNSINFLONSGID		
Oy	270	SYOPGIRVLPESHSDVDOLKDKDESANIT		
Db	756	OLASAKVFRKIPREINLDFAKHEASIKSASVIVISLAVYDELFGNSKSKILLERKHIDS		
Oy	315	FR		
Db	816	FMWYILLICIGSHLLIGFCVAVAGFTKNGRAKAYINLTYLFGKDLDEFFERRYS		
Oy	338	TDHOYLERIPICA		
Db	874	LDRQDKNTLLDAVEIKNTYWKAKVAGSYIYAYNK		
Oy	391	VPHGDDIIFOCDAKTRIPDVKALONTSLDYSKITWITTKHOKGPKVAVVH		
Db	531	LIEGADILL		

Search completed: April 3, 2002, 17:10:45
Job time: 245 sec

XX Novel nucleic acids and polypeptides, useful for treating disorders
PI such as central nervous system injuries -

XX Example 2: SEQ ID NO 5842: 10078pp: English.

CC The invention relates to human nucleic acids (AA157788-AA151369) and
CC the encoded polypeptides (AA158642-AA142213) with nucleotide,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening and
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 307 AA:

Query Match 33.7%: Score 922: DB 22: Length 307:
Best Local Similarity 99.4%: Pred. No. 2.5e-86:
Matches 174: Conservative 1: Mismatches 0: Indels 0: Gaps 0:

OY 328 PCCQSAATLITLDHYLFPLCAILKOKAPOYRIRAKLRSTPRRLTOSVLCRCPKCHL 387
ID ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AC 1 rcqgisatclitdhgylerplcalikqkqgqyrlraklrstprllqgsxhhpcpchl 50

OY 388 LOEVHEGCDLIIIFODGAKITPDVKLONTSLVDSKIKITKNOKGRKVAVHEFKNGIILPL 447
ID ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 61 lgerpheddlitlfdgatkcpdvklqntsljydsakiwtknqkykvavhfrkmqilpl 120

OY 448 SNECLLEGGTLESTICKLSNKFNSVIVPSGHEDELELDLSAPPLIOGTIHNYG 502
ID ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 121 snecllllegtlstclcklsnkfnsvlprsghelelldlsaplllqgtvhyg 175

RESULT 6

AA151264
ID AA151264 standard: Protein: 30 AA.

AC AA151264:

AC 17-OCT-2001 (first entry)

DE Peptide #8301 encoded by probe for measuring placental gene expression.

KW Probe: microarray: human; Placenta: antenatal diagnosis;
KW genetic disorder.

OS Homo sapiens.

FN WO200157272-A2.

PD 09-AUG-2001.

FE 30-JAN-2001: 2001WO-US00653.

FR 04-FEB-2000: 2000US-0180312.

FR 26-MAY-2000: 2000US-0207455.

PR 30-JUN-2000: 2000US-0608408.

PR 03-AUG-2000: 2000US-0632356.

FR 21-SEP-2000: 2000US-0234587.

FR 27-SEP-2000: 2000US-0236359.

FR 04-OCT-2000: 2000GB-0024253.

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI: 2001-488897/53.

PI Human genome-derived single exon nucleic acid probes useful for
PI analyzing gene expression in human placenta -

PS Claim 27: SEQ ID NO 34533: 654pp: English.

CC The present invention relates to single exon nucleic acid probes (SENP:
CC see AA131315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.

XX Sequence 30 AA:

Query Match 5.7%: Score 156: DB 22: Length 30:
Best Local Similarity 100.0%: Pred. No. 5.4e-09:
Matches 30: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 389 OFVHEGSDLIIFODGAKITPDVKLONTSL 418
ID ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1 lgerpheddlitlfdgatkcpdvklqntsl 30

RESULT 7
AAG58505
ID AAG58505 standard: Protein: 467 AA.

AC AAG58505:

AC 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 75533.

KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay: genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

FN EP1033405-A2.

PD 06-SEP-2000.

FE 25-FEB-2000: 2000EP-0301439.

FR 25-FEB-1993: 99US-0121825.

FR 05-MAR-1999: 99US-0123180.

FR 05-MAR-1999: 99US-0123548.

FR 23-MAR-1999: 99US-0125788.

FR 25-MAR-1999: 99US-0126254.

FR 29-MAR-1999: 99US-0126785.

FR 01-APR-1999: 99US-0127452.

FR 06-APR-1999: 99US-0128234.

FR 08-APR-1999: 99US-0128714.

FR 15-APR-1999: 99US-0129845.

FR 19-APR-1999: 99US-0130077.

FR 21-APR-1999: 99US-0130449.

FR 23-APR-1999: 99US-0130510.

FR 23-APR-1999: 99US-0130891.

FR 28-APR-1999: 99US-0131449.

FR 30-APR-1999: 99US-0132048.

FR 30-APR-1999: 99US-0132407.

FR 04-MAY-1999: 99US-0132484.

FR 05-MAY-1999: 99US-0132485.

FR 05-MAY-1999: 99US-0132485.

FR 07-MAY-1999: 99US-0132487.

FR 07-MAY-1999: 99US-0132863.

PR 11-MAY-1999: 99US-0134256.
PR 14-MAY-1999: 99US-0134218.
PR 14-MAY-1999: 99US-0134219.
PR 14-MAY-1999: 99US-0134221.
PR 14-MAY-1999: 99US-0134270.
PR 18-MAY-1999: 99US-0134768.
PR 19-MAY-1999: 99US-0134941.
PR 20-MAY-1999: 99US-0135124.
PR 21-MAY-1999: 99US-0135293.
PR 24-MAY-1999: 99US-0135629.
PR 25-MAY-1999: 99US-0136021.
PR 27-MAY-1999: 99US-0136392.
PR 28-MAY-1999: 99US-0136782.
PR 01-JUN-1999: 99US-0137222.
PR 03-JUN-1999: 99US-0137528.
PR 04-JUN-1999: 99US-0137562.
PR 07-JUN-1999: 99US-0137724.
PR 08-JUN-1999: 99US-0138094.
PR 10-JUN-1999: 99US-0138540.
PR 14-JUN-1999: 99US-0138847.
PR 16-JUN-1999: 99US-0139119.
PR 16-JUN-1999: 99US-0139452.
PR 16-JUN-1999: 99US-0139453.
PR 17-JUN-1999: 99US-0139452.
PR 17-JUN-1999: 99US-0139452.
PR 18-JUN-1999: 99US-0139454.
PR 18-JUN-1999: 99US-0139455.
PR 18-JUN-1999: 99US-0139456.
PR 18-JUN-1999: 99US-0139457.
PR 18-JUN-1999: 99US-0139458.
PR 18-JUN-1999: 99US-0139459.
PR 18-JUN-1999: 99US-0139460.
PR 18-JUN-1999: 99US-0139461.
PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139753.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139899.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140695.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142830.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0143542.
PR 13-JUL-1999: 99US-0143524.
PR 14-JUL-1999: 99US-0144005.
PR 15-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 16-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144352.
PR 20-JUL-1999: 99US-0144632.
PR 20-JUL-1999: 99US-0144884.
PR 21-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145086.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145099.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.

PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145918.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 03-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 05-AUG-1999: 99US-0147302.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148655.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 20-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 23-AUG-1999: 99US-0150566.
PR 25-AUG-1999: 99US-0150884.
PR 26-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151080.
PR 27-AUG-1999: 99US-0151303.
PR 30-AUG-1999: 99US-0151438.
PR 31-AUG-1999: 99US-0151930.
PR 01-SEP-1999: 99US-0152363.
PR 07-SEP-1999: 99US-0153070.
PR 10-SEP-1999: 99US-0153758.
PR 13-SEP-1999: 99US-0154039.
PR 15-SEP-1999: 99US-0154039.
PR 16-SEP-1999: 99US-0154779.
PR 20-SEP-1999: 99US-0155139.
PR 22-SEP-1999: 99US-0155486.
PR 23-SEP-1999: 99US-0155659.
PR 24-SEP-1999: 99US-0156458.
PR 28-SEP-1999: 99US-0156596.
PR 29-SEP-1999: 99US-0157117.
PR 04-OCT-1999: 99US-0157753.
PR 05-OCT-1999: 99US-0157865.
PR 06-OCT-1999: 99US-0158029.
PR 07-OCT-1999: 99US-0158232.
PR 08-OCT-1999: 99US-0158369.
PR 12-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 13-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 14-OCT-1999: 99US-0159684.
PR 18-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.

```

PR 22-OCT-1999: 990S-0160981.
PR 22-OCT-1999: 990S-0160989.
PR 25-OCT-1999: 990S-0161404.
PR 25-OCT-1999: 990S-0161405.
PR 25-OCT-1999: 990S-0161406.
PR 26-OCT-1999: 990S-0161359.
PR 26-OCT-1999: 990S-0161360.
PR 26-OCT-1999: 990S-0161361.
PR 28-OCT-1999: 990S-0161920.
PR 28-OCT-1999: 990S-0161992.
PR 28-OCT-1999: 990S-0161993.
PR 29-OCT-1999: 990S-0162142.

```

Query Match 4.6% Score 125 DB 21 Length 457
 Best Local Similarity 24.0% Pred. No. 0.0009
 Matches 65: Conservative 38: Mismatches 92: Indels 76: Gaps 13:

```

OY 23 VAVVGVVFFKRPPLSKGIDYCSVYIVDQT--NKKLTCLLSCGNALPILYKKGSDIVR 80
DB 24 VSLIGIVLEGREPQCRMDWICILIRIDDTYSPGLIVVNVISKLEGLPQIKNDMMIL 83
OY 61 FHRLEIQVYK--KETOGIIS---SGFASLTFEGILGAPILPRISKEYFTIEDHK---- 131
DB 64 FRIKMGITDSQKRVNAACSRVSSFA--LLEG-----VDFCYGCSGTFNEEEAL 132
OY 132 ---MYEALPVNA-----STHNSP-SWTLKLCIDVQPMOYFDLTQQLGKALV 174
DB 133 YKSAAMDILKVFACGSGVLIKAMQSLYRTKPCSEVFSILREIKIGRFDLVCRIIHAD 192
OY 175 DGASFLKRWGDIRTP-----FPSWRV---LIODLVLE-----GDLSIHHL 213
DB 193 TSEVLT--VWDGTDAPPASILAKRSEEDKAFSSISVHLISITDVLISFTVGLIRVHS 249
OY 214 QMLITIDILVYDNHVAHVASLKVSGFLRIYSL 244
DB 250 S-----HLFYRAKPGDWKLYHL 267

```

RESULT 8

AAR55694

ID AAR55694 standard: Protein: 2391 AA.

AC AAR55694:

DI 06-DEC-1994 (first entry)

XX Carbamoyl-phosphate-synthetase II.

XX Carbamoyl-phosphate-synthetase II: CPSII: pscpsii gene:

XX malarial.

XX Plasmodium falciparum.

OS Plasmodium falciparum.

XX Key Location/Qualifiers

FH Domain 1..590 /note= "glutamine:amidotransferase domain"

FI Domain 1..270 /note= "structural subdomain"

FI Peptide 271..482 /note= "insert sequence"

FI Domain 483..690 /note= "glutaminase subdomain"

FI Domain 691..2391 /note= "carbamoyl-phosphate-synthase domain"

FI Domain 691..1254 /note= "ATP binding subdomain CPSa"

FI Peptide 1255..1857 /note= "insert sequence"

FI Domain 1858..2391 /note= "ATP binding subdomain CPSb"

XX WO9412643-A.

```

PD 09-JUN-1994.
XX 02-DEC-1993: 93WO-A000617.
PF 03-DEC-1992: 92AU-0006206.
PR 15-DEC-1992: 92AU-0006380.
XX (UNIX ) UNISEARCH LTD.
XX Flores MV, O Sullivan WJ, Stewart TS:
XX WPI: 1994-200271/24.
XX N-PSDB: AA062924.
PI Nucleic acid encoding carbamoyl phosphate synthetase II -
PI isolated from Plasmodium falciparum, used to develop prods. for
PI the treatment of malaria.

```

Disclosure: Page 6-16; 31pp: English.

The cDNA sequence encoding the carbamoyl-phosphate-transferase II (CPSII) of Plasmodium falciparum was determined. The cDNA encodes a protein that includes 2 insert sequences not found in other CPSII proteins. The first separates the putative structural subdomain and the glutaminase subdomain of the glutamine:amidotransferase subunit of CPSII, while the second separates 2 ATP binding subdomains of the CPSII subunit, CPSa and CPSb.

Sequence 2391 AA:

Query Match 3.9% Score 106 DB 15 Length 2391
 Best Local Similarity 20.9% Pred. No. 1.3
 Matches 109: Conservative 80: Mismatches 230: Indels 102: Gaps 25:

```

OY 7 TWVITPLNQ--LKGGTIVNVYGV--VKFFKRPPLSKGIDYCSVYIVDQTNWKL-CLL 51
DB 939 SNLEYKFKREIAIK:ITLNIIGECNIQGLNPQTG---EYCI-----ENARISRSA 990
OY 62 FSGNVEALPIIYKNGDIVFHH--RLKIOVYKKEGOGITSSGFASLTFEGILG--APIP 116
DB 991 IASKATGYPLAYISAKIAGYDLISLKNLSLKKT-----ACTEPSIDYITKIP 1040
OY 117 RISSKYFNFT----EDKKHVEALRVNASTHNSPFWTLKLCIDVQPMOYFDLTQQLGKA 172
DB 1041 RVDLHKFEFASNTMNSMKSVGEVMSIGRTFEESIKSLRCDIDNYIGFNTYC----- 1094
OY 173 EYDGASFLKRWGDIRTPFPSPWRV--LIODLVLEGDLSIHRLQMLITIDILVYDNHVA 230
DB 1095 -IDKGE--KKICEELKNPSK-RIDAIHQAFILNPMKIKHETLHDYWL----- 1141
OY 231 RELKVGSLRIYSLHTKLOSANSENQTMLSDEFHLHGSTYGRGIRVLPESNSVDVQLK 290
DB 1142 -----HKFYNIYNLGNKIKLKLLEGISTFNDIKYFKKHGFSQGIHYLSFNSDNNNN 1196
OY 291 DLFSAHLTANQHSVYICQSEPDSPSSGSVSLYEVRQQLSATL--LTQHGYL----- 343
DB 1197 NLSGCRVEN--DVMKYREKIGLIP-----HKYIDLTLSAEFPALNTYLYTYGQG 1245
OY 344 ERTPLCAILKOKAPOOYRIAPAKLSYKPRRLPQSYAKLHCPKCHLQVPHNEGDLIIFOD 403
DB 1245 ENDVYIPLMKTK-----KICLLNKNRANKKVVHV-KHLYNEVVDKIDGJHKN 1295
OY 404 GAKTIPDVKLQNTSLYDSKIMTTKNQGRKVVHNVKNGNLPLPSNECLLLEGITLSEI 453
DB 1296 NNNN-----MNSGNVEKCKLNKESYGNNSNGLNINLNIENIC-----HDSIOKN 1346
OY 454 CK--LSNKFNSVIPVRSCHDELLELDLSAPFLIOGTIIHYG 502
DB 1347 IK::LINSNLSI-----SNNEVE-LNINCEVERAGSHIYIG 1302

```

RESULT 9

AAW05556
 ID AAW05556 standard: Protein: 655 AA.
 XX
 AC AAW05556;
 XX
 DI 12-FEB-1997 (first entry)
 XX
 DE Serine kinase SRPK1.
 XX
 KW Human: serine kinase; SRPK1: SR domain: SR family: splicing factor;
 KW cell proliferation; malignancy; psoriasis; pemphigus vulgaris;
 KW Behcet's syndrome; acute respiratory distress syndrome; leukemia;
 KW rheumatoid arthritis; AIDS; vasculitis; septic shock;
 KW acute inflammation; lipid histiocytosis; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 11..21 "Potential nuclear localisation signal"
 FT 80..100 "Conserved kinase domain I"
 FT Misc-difference 87 /note= "Conserved residue"
 FT Misc-difference 89 /note= "Conserved residue"
 FT Misc-difference 94 /note= "Conserved residue"
 FT Misc-difference 103..119 /note= "Conserved residue"
 FT Domain 109 /note= "Conserved kinase domain II"
 FT Misc-difference 122..129 /note= "Conserved residue"
 FT Domain 132..146 /note= "Conserved kinase domain III"
 FT Domain 148..169 /note= "Conserved kinase domain IV"
 FT Domain 172..205 /note= "Conserved kinase domain V"
 FT Domain 208..225 /note= "Conserved kinase domain VI"
 FT Domain 213 /note= "Conserved kinase domain VII"
 FT Misc-difference 215 /note= "Conserved residue"
 FT Misc-difference 218 /note= "Conserved residue"
 FT Misc-difference 257..277 /note= "Conserved residue"
 FT Peptide 297..307 /note= "Potential nuclear localisation signal"
 FT Domain 492..507 /note= "Conserved kinase domain VIII"
 FT Misc-difference 497 /note= "Conserved residue"
 FT Misc-difference 499 /note= "Conserved residue"
 FT Misc-difference 509..523 /note= "Conserved residue"
 FT Domain 520 /note= "Conserved kinase domain VIII"
 FT Misc-difference 520 /note= "Conserved residue"
 FT Misc-difference 521 /note= "Conserved residue"
 FT Domain 526..553 /note= "Conserved kinase domain IX"
 FT Misc-difference 533 /note= "Conserved residue"
 FT Misc-difference 538 /note= "Conserved residue"
 FT Domain 554..573 /note= "Conserved residue"
 FT Domain 521..554 /note= "Conserved kinase domain X"
 FT Domain 541 /note= "Conserved kinase domain XI"
 FT Misc-difference 641

FT /note= "Conserved residue"

XX US5559019-A.
 XX 24-SEP-1996.
 XX 22-JUN-1994: 94US-0264002.
 XX 22-JUN-1994: 94US-0264002.
 XX (REGC) UNIV CALIFORNIA.
 XX Fu X, Gui J:
 XX WPL 1996-442410/44.
 XX N-PSDR: AAT43897.
 XX
 PI New isolated serine kinase nucleic acid encoding SRPK1 - used to
 PI develop prods. which can be used in detection and therapy, partic.
 PI for cell proliferative disorders
 XX
 PS Claim 2: Fig 2: 42pp: English.
 XX
 CC This sequence represents human serine kinase SRPK1. In the SRPK1
 CC DNA sequence, the sequence preceding the ATG initiation codon contains
 CC stop codons in all three reading frames. The sequence surrounding the
 CC initiation codon conforms to the kozak consensus sequence. SRPK1 acts
 CC to phosphorylate serine residues of the SR domain of the SR family of
 CC splicing factors. The phosphorylation of these sites by SRPK1
 CC affects the ability of the splicing factors to mediate pre-mRNA
 CC splicing and regulates the cellular distribution of splicing factors
 CC during the cell cycle. SRPK1 has a mol. wt. of 92 kD as determined
 CC by SDS-PAGE. The modulation of SRPK1 activity can be used to treat
 CC cell proliferative disorders, e.g. malignancies, psoriasis, pemphigus
 CC vulgaris, Behcet's syndrome, acute respiratory distress syndrome,
 CC leukemia, rheumatoid arthritis, AIDS, vasculitis, septic shock and
 CC other types of acute inflammation, and lipid histiocytosis. The
 CC sequence given in fig. 2 is longer than the sequence given in the
 CC sequence listing in the body of the specification.
 XX
 SO Sequence 655 AA:

Query Match 3.8%; Score 105; DB 17; Length 655;
 Best Local Similarity 20.9%; Pred. No. 0.19; 98; Indels 94; Gaps 13;
 Matches 61; Conservative 39; Mismatches 94; Gaps 13;

QY 247 KLGSMNSENTMLSLFHLHGTS-----YGRQIRVLPESNSDVQDKKDLSEAN---- 296
 DB 329 KLEESTLIGDQTLMERDELGGAELINCNGVIEVINYLTGSMNELIRKEDLHMANDCV 388
 QY 297 LITNQHSDVVICOSPPDPSFSSGVSILYVERCOQLSATILTHOYLERPLC---ALIK 353
 DB 389 qnlqessfl-----slpqdsslsqetdsclplseve-----sdmwcqssstly 434
 QY 354 OKAPQGYRIIRAKLRKYPRRLFOVKLHCPCXCHLDQVPHFGDLIIFOGATIKTPDKVL 413
 DB 435 qsfesqh-----lsqqlgestraeip-edegeqegmnpd----- 458
 QY 414 CNTSLYDSKIWTRNCKGRKVAIVFVKNNGILP--SNECLLLIEGTLSEIKLSKEN 471
 DB 459 -----nkxstlaqnlfcn---plepknaeklxkialdiqnaecvnhht 509
 QY 472 SVLPVPSGHEDELLDLAPFLIOGTIHHVGTGCTP-----PICVCYDYLYT 518
 DB 510 editlrq-yrleel-----lsqyntpdpadwstacmefelat 546
 RESULT 10
 AAY58113
 ID AAY58113 standard: Protein: 1112 AA.
 XX
 AC AAY58113;

```

XX 07-MAR-2000 (first entry)
XX
XX Haemorrhagic enteritis virus (HEV) RNA polymerase.
XX
XX Haemorrhagic enteritis virus (HEV). Immune system: Turkey: abn. no. 1000.
XX
XX Inoculation: haemorrhagic. Immunosuppression: RNA polymerase.
XX
XX Rubella vaccine: antiviral: recombinant: vector: gene therapy.
XX
XX Haemorrhagic enteritis virus.
XX
XX MO9960131-A2.
XX
XX 25-NOV-1999.
XX
XX 19-MAY-1999: 99WO-11.0026R.
XX
XX 20-MAY-1998: 98IL-0124557.
XX
XX (ABIC ) ABIC LTD.
XX
XX Ptkovskij J, Maalem M, Rej Koren Z, Kispel S, Shmueli I;
XX Peretz Y, Gutter B, Gallili G, Michael A, Goldberg D;
XX
XX WPI: 2000-062458/05.
XX
XX N-PSDB: AA245358.
XX
XX New hemorrhagic enteritis virus genes useful as vaccines for treating
XX viral infection in domesticated birds e.g. turkey and in humans -
XX
XX Claim 31: Page 63-67: 89pp: English.
XX
XX Sequences AAY58101-Y58115 represent proteins encoded by the
XX haemorrhagic enteritis virus (HEV). HEV is an adenovirus which infects
XX turkeys, causing intestinal haemorrhaging and immunosuppression.
XX The structural proteins include the 97 kD hexon protein (AAY58107),
XX which is an outer capsid monomer; the 50 kD penton base protein
XX (AAY58103); and the fibre protein (AAY58111), which anchors the penton
XX base protein and plays an important role in the first attachment of the
XX virus to the cell receptor. Other HEV proteins are 52k (AAY58101), 111a
XX (AAY58102), core protein I (CPI, AAY58104), core protein II (CPII,
XX AAY58105), pVI (AAY58106), endoprotease (EP, AAY58108), 100k
XX (AAY58109), pVII (AAY58110), Iva2 (AAY58112), DNA polymerase (POL,
XX AAY58113), precursor terminal protein (PTP, AAY58114), and DNA binding
XX protein (DBP, AAY58115). These proteins, or the genes encoding them, may
XX be used in the preparation of a subunit vaccine against the virus. Such
XX vaccines are likely to be effective, and also inexpensive, making their
XX use economically worthwhile. Additionally, the fibre protein, which
XX mediates viral attachment, may be modified to alter its host cell
XX specificity. A recombinant HEV may be constructed for use as a vector for
XX gene therapy. The nucleotides encoding the proteins may also be used for
XX diagnostic purposes, or may be used as a source of primers and probes.
XX
XX Sequence 1112 AA:
XX
XX
XX

```

Query Match 3 8%: Score 104; DB 21: Length 1112;
 Best Local Similarity 19.0%: Pred. No. 0.58;
 Matches 96: Conservative 78: Mismatches 171: Indels 160: Gaps 25;

```

QY 7 TNYIYPLNQLK-----GGTIVVYGVVFFKPYLSKGDYCSVVIYDQINWK 56
   : 1111 1: : 11: 1: : 1: 1: : 1: 1: :
DB 2 skyiytlqneipyrjgickkdkatijefawynlfncckkial-----kvkagqsyj 53
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
QY 57 LTCLFSGNEVALPIIYKGGDIVRFHRL--KIOYKKEKTOGI-----TSSGASLTF 105
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
DB 54 llsckykpc-----nsdienvknlpvksiltlwkklgtgcmqmetkshyefield 104
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
QY 107 EGTLAGAPLIPPTSSKYFFNTTIEDHKNVEALPYWASIHMSPSNLLKLC-DVDPMPQYFD-- 153
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
LH 105 -----lydsksklcwksctgqgkcsseqrltysaih-----tcndlrstyymsv 149
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:

```

```

QY 164 -----ETGOLIGKAEVMAFLKLVMD-GTRTFPSMRVLIDPLV-----ESGLS 208
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
DB 149 dndtktksetipqgqivmmentkklil--lydfof fllfoosvltmptlltstfnd 204
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
QY 209 HHHHGLGHLIDLLVYDNIHVAKSLKVGSPFLKIVSLHFKLQSNSENF-----ML 259
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
DB 206 -----dellctcktdla-----kkslqat kstlylwkswkvl 248
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
QY 260 SEIHH-----LHRSKSGKGLKVLSPNSNDPLV--KKDLSS-ANLJANSHVLCOSE 310
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
DB 249 seetkklrmltlatllyfkliltpenrelledaqgqlsestlnlnstehnlmslk 298
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
QY 311 PHHSFPSSGSVSLYFVERCOO-----LSATLLTDHLYERTPLCALIKOKAPDOGYRIAR 365
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
DB 299 vkpklf-----loyltqlmngsfndllatqlldendldplllvernfmpgrgl--- 350
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
QY 366 LSVYKRPPLFGSVKLMCP-----KCLILQF-----PHRRDLDIIFODGATK 407
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
DB 351 -----lndlttkfmpdyvayakeekknkksellsqikealpfheslknlylksmvr 402
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
QY 408 TTIVKLIQHTSLYD-SKIWTTKNOKG 431
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:
DB 403 -dlldltstskvknkaynkklhkg 426
   : 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1: 1:

RESULT 11
AAB31889
ID AAB31889 standard: Protein: 4393 AA.
AC AAB31889:
XX
XX 15-MAY-2001 (first entry)
XX
XX Amino acid sequence of a human protein.
XX
XX DE
XX
XX Human: perlecan: retinol-binding plasma protein: calgranulin B: vaccine:
XX ganglioside GM2 activator: saposin B: degenerative disease: glial cell:
XX neurological disease: auto-immune disease: multiple sclerosis: toxicity:
XX Alzheimer's disease: Parkinson's disease: amyotrophic lateral sclerosis:
XX rheumatoid polyarthritis: lupus erythematosus: gene therapy.
XX
XX HOMO sapiens.
XX
XX W0200105442-A2.
XX
XX 25-JAN-2001.
XX
XX 17-JUL-2000: 2000WO-FR02057.
XX
XX 15-JUL-1999: 99FR-0009372.
XX
XX (IMMR ) BIOMERIEUX STELHYS.
XX
XX Foecklin D, Kolbe H, Charles M, Marcus C, Santoro L, Perron H;
XX
XX WPI: 2001-159475/15.
XX
XX Detecting, preventing and treating degenerative, neurological and
XX autoimmune diseases, particularly multiple sclerosis, using specified
XX polypeptides or related nucleic acid or ligand -
XX
XX Claim 1: Page 138-152: 209pp: French.
XX
XX The present sequence represents a human protein, which is used in the
XX method of the invention. The specification describes a method which uses
XX at least one polypeptide or polynucleotide sequence belonging to the
XX perlecan, precursor of the retinol-binding plasma protein, precursor of
XX the ganglioside GM2 activator, calgranulin B or saposin B protein
XX families. The method is used for detecting, preventing or treating a
XX degenerative, neurological and/or auto-immune disease. The
XX polynucleotides and polypeptides are used for diagnosis, prognosis,
XX prevention and treatment of multiple sclerosis (in its various forms
XX and phases). They may also be useful in cases of e.g. Alzheimer's and

```


CC Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
 CC polyarthritis and lupus erythematosus, including use as vaccines and
 CC in gene therapy (expression of sense or antisense sequences). They can
 CC also be used to assess efficacy of potential therapeutic agents, and
 CC particularly compounds that reduce or inhibit toxicity towards glial
 CC cells.

XX Sequence 4393 AA:

Query Match 3.7% Score 102; DB 22; Length 4393;
 Best Local Similarity 24.1% Pred. No. 9.6;
 Matches 81; Conservative 38; Mismatches 137; Indels 8; Gaps 13;

OY 48 TIVDQINWKLTLFSGNVEALPIIYKNGD-----VFHRLKI-----QVYKKE 92
 DB 2257 tvaegtlldscv-agahagvtykrgslparhagvgrlylfqspadagqvrra 2315
 OY 93 TCGTSSGFASLTFEGTGA-----PI-IPRTSKTFNFTED----- 129
 DB 2316 snqmeas--ltvtytgqganlaypagstqplilepsssgyaegqtlldncvvpqgshag 2373
 OY 130 ---HKMEALRWKASTHMSPTLLKLCDOVPMQYFDLTCLGKAEDGASFLKXWDG 186
 DB 2374 vtwkrggsjpxvthqths-----llrlqspadagvtykrgslpar 2429
 OY 147 TRTP----FPCWFLDQVLLEGDLSHIRLQNTLIDLVYDN-HVHV-----AR 231
 DB 2430 gsvpalgvtflpvtl-----essssgyaegqtlldncvqgahagvtykrggsllpar 2482
 OY 232 SLKVGSEFLI-----YSLHKLOSMSSENOTMSTLEFHLHGCTSYGPGIRVLEES 281
 DB 2483 hqthgsstlrllyqtpadagvtykrgssqgsvltqlqrlsqshsqvayprvrls 2542
 OY 282 NSDVLQKDELSANLTANQSHVDYICQEPDSEPS 317
 DB 2543 ssaslaqghltdlncvqgahagvtykrggsllps 2578

RESULT 12

AAV34460
 ID AAV34460 standard; Protein: 684 AA.

XX AAV34460:

XX 25-AUG-1993 (first entry)

XX Porphyromonas gingivalis protein PG116.

XX Porphyromonas gingivalis; PG: periodontal disease; gingivitis;

XX vaccine; antigenic.

XX Porphyromonas gingivalis.

XX WO9929870-A1.

XX 17-JUN-1999.

XX 10-DEC-1998: 98MO-AUD1023.

XX 04-AUG-1998: 98AU-0005028.

XX 10-DEC-1997: 97AU-0000839.

XX 31-DEC-1997: 97AU-0001182.

XX 30-JAN-1998: 98AU-0001546.

XX 10-MAR-1998: 98AU-0002264.

XX 09-APR-1998: 98AU-0002911.

XX 23-APR-1998: 98AU-0003128.

XX 05-MAY-1998: 98AU-0003338.

XX 22-MAY-1998: 98AU-0003554.

XX 29-JUL-1998: 98AU-0004917.

FI Aajus CI, Barr IG, Hocking DM, Margolis MB, Patterson MA;

FI Posn BC, Polhel LJ, Webb EA;

XX WPI: 1999-385613/32.

XX N-PSDB: AAX91578.

XX Antigenic Porphyromonas gingivalis peptides for preventing

XX gingivitis

XX Claim 1: Page 438-439; 588pp; English.

CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAV34318 to
 CC AAV34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.

XX Sequence 684 AA:

Query Match 3.6% Score 99; DB 20; Length 684;
 Best Local Similarity 21.1% Pred. No. 0.83;
 Matches 101; Conservative 69; Mismatches 177; Indels 132; Gaps 26;

OY 57 LTCLFSGNVEALPIIYKNGDVFHRLKIQVY-----KKFQGITSSGFASLTFEGTL 110
 DB 270 ltqilly-----fplklnmndi---qtnkqilycngyvltdevqglvpd-fllt-lhavt 318
 OY 111 GAPLIPRTSKRYFNFTEDHKKMEALRWKASTHMS-----PSWTL 151
 DB 319 dsdpdlnvrsryqsdanvkk-----sshtlkvvadrlleifkndrptfeekwds 371
 OY 152 KLCDOVPMQYFDLTCLGKAEDGASFLKXWDGTRTPPSRVLIQVLE----- 204
 DB 372 klf---veygmldexf--yeraakflltdmgnhytldetvrltvegvqtdkqgvvy 425
 OY 205 -----GDLSHIRLQNTLIDLVYDNHV--HVA-----SLKVGSEFLIYFS----- 243
 DB 426 lyaekbhqgyswhkrasdkysvmjldgqldphivellqekthfvrvdedtlmnlr 485
 OY 244 --LHFKLOSMSSENOTMSTLEFHLHGCTSYGPGIRVLPESNSVDOLKDLDSA--NLTA 299
 DB 485 keeraecklsdteratvkl-fear-----lprde-----khhfvatesjga 527
 OY 300 NQSHVDVICOSE-----PDDSPSSGSVSLYEVEPCOOLSATILTDHOYLERTPLCALIKOK 355
 DB 528 egeaillltgaefmrmtmdaglqpgmsfyg-elipdsynlvintdpltdrv-lsgekesv 585
 OY 356 APOGYRIPAKLRSTYKR--RLFOSVKLHCPKCHLLOEVPHEGDLIDIFODGATKTPDVKL 413
 DB 585 epslteltraklaekkaeaklldckkbp-----eeip-----vatkeak- 626
 OY 414 QNTSLYDSKIKWITKQKGRVAVHFNKNGILPLSNCLL--LIEGSLSECKLSNK 459
 DB 627 ennaevagakegsgindqltkya---qdnelllqgldalllqsglltgeaiaeftrrsqr 682

RESULT 13

AAV34336
 ID AAV34336 standard; Protein: 686 AA.

XX AAV34336:

XX 25-AUG-1999 (first entry)

XX Porphyromonas gingivalis protein PG116.

XX Porphyromonas gingivalis; PG: periodontal disease; gingivitis;

XX vaccine; antigenic.

PR 10-JUN-1999: 99US-0138847-
PR 14-JUN-1999: 99US-0139119-
PR 16-JUN-1999: 99US-0139452-
PR 16-JUN-1999: 99US-0139453-
PR 17-JUN-1999: 99US-0139492-
PR 18-JUN-1999: 99US-0139454-
PR 18-JUN-1999: 99US-0139455-
PR 18-JUN-1999: 99US-0139456-
PR 18-JUN-1999: 99US-0139457-
PR 18-JUN-1999: 99US-0139458-
PR 18-JUN-1999: 99US-0139459-
PR 18-JUN-1999: 99US-0139460-
PR 18-JUN-1999: 99US-0139461-
PR 18-JUN-1999: 99US-0139462-
PR 18-JUN-1999: 99US-0139463-
PR 18-JUN-1999: 99US-0139750-
PR 18-JUN-1999: 99US-0139753-
PR 21-JUN-1999: 99US-0139817-
PR 22-JUN-1999: 99US-0139899-
PR 23-JUN-1999: 99US-0140353-
PR 23-JUN-1999: 99US-0140354-
PR 24-JUN-1999: 99US-0140695-
PR 28-JUN-1999: 99US-0140823-
PR 29-JUN-1999: 99US-0140991-
PR 30-JUN-1999: 99US-0141287-
PR 01-JUL-1999: 99US-0141682-
PR 01-JUL-1999: 99US-0142154-
PR 02-JUL-1999: 99US-0142055-
PR 06-JUL-1999: 99US-0142390-
PR 08-JUL-1999: 99US-0142803-
PR 09-JUL-1999: 99US-0142920-
PR 12-JUL-1999: 99US-0142977-
PR 13-JUL-1999: 99US-0143542-
PR 14-JUL-1999: 99US-0143624-
PR 15-JUL-1999: 99US-0144005-
PR 16-JUL-1999: 99US-0144086-
PR 16-JUL-1999: 99US-0144325-
PR 19-JUL-1999: 99US-0144331-
PR 19-JUL-1999: 99US-0144332-
PR 19-JUL-1999: 99US-0144333-
PR 19-JUL-1999: 99US-0144334-
PR 19-JUL-1999: 99US-0144335-
PR 19-JUL-1999: 99US-0144352-
PR 20-JUL-1999: 99US-0144532-
PR 20-JUL-1999: 99US-0144884-
PR 21-JUL-1999: 99US-0144814-
PR 21-JUL-1999: 99US-0145086-
PR 21-JUL-1999: 99US-0145088-
PR 22-JUL-1999: 99US-0145089-
PR 22-JUL-1999: 99US-0145089-
PR 22-JUL-1999: 99US-0145089-
PR 22-JUL-1999: 99US-0145089-
PR 22-JUL-1999: 99US-0145192-
PR 23-JUL-1999: 99US-0145145-
PR 23-JUL-1999: 99US-0145218-
PR 23-JUL-1999: 99US-0145224-
PR 26-JUL-1999: 99US-0145276-
PR 27-JUL-1999: 99US-0145911-
PR 27-JUL-1999: 99US-0145912-
PR 27-JUL-1999: 99US-0145919-
PR 28-JUL-1999: 99US-0145951-
PR 02-AUG-1999: 99US-0146386-
PR 02-AUG-1999: 99US-0146388-
PR 02-AUG-1999: 99US-0146389-
PR 03-AUG-1999: 99US-0147038-
PR 04-AUG-1999: 99US-0147204-
PR 05-AUG-1999: 99US-0147302-
PR 05-AUG-1999: 99US-0147250-
PR 06-AUG-1999: 99US-0147303-
PR 06-AUG-1999: 99US-0147415-
PR 09-AUG-1999: 99US-0147493-
PR 09-AUG-1999: 99US-0147935-

PR 10-AUG-1999: 99US-0148171-
PR 11-AUG-1999: 99US-0148319-
PR 12-AUG-1999: 99US-0148341-
PR 13-AUG-1999: 99US-0148565-
PR 13-AUG-1999: 99US-0148684-
PR 16-AUG-1999: 99US-0149368-
PR 17-AUG-1999: 99US-0149175-
PR 18-AUG-1999: 99US-0149426-
PR 20-AUG-1999: 99US-0149723-
PR 20-AUG-1999: 99US-0149929-
PR 23-AUG-1999: 99US-0149902-
PR 25-AUG-1999: 99US-0149930-
PR 25-AUG-1999: 99US-0150084-
PR 26-AUG-1999: 99US-0151065-
PR 27-AUG-1999: 99US-0151066-
PR 27-AUG-1999: 99US-0151080-
PR 30-AUG-1999: 99US-0151303-
PR 31-AUG-1999: 99US-0151330-
PR 01-SEP-1999: 99US-0152363-
PR 07-SEP-1999: 99US-0153070-
PR 10-SEP-1999: 99US-0153758-
PR 13-SEP-1999: 99US-0154018-
PR 15-SEP-1999: 99US-0154039-
PR 16-SEP-1999: 99US-0154779-
PR 20-SEP-1999: 99US-0155139-
PR 22-SEP-1999: 99US-0155486-
PR 23-SEP-1999: 99US-0155659-
PR 24-SEP-1999: 99US-0156458-
PR 28-SEP-1999: 99US-0156596-
PR 29-SEP-1999: 99US-0157117-
PR 04-OCT-1999: 99US-0157753-
PR 05-OCT-1999: 99US-0157865-
PR 06-OCT-1999: 99US-0158029-
PR 07-OCT-1999: 99US-0158232-
PR 08-OCT-1999: 99US-0158369-
PR 12-OCT-1999: 99US-0159293-
PR 13-OCT-1999: 99US-0159294-
PR 13-OCT-1999: 99US-0159295-
PR 14-OCT-1999: 99US-0159329-
PR 14-OCT-1999: 99US-0159330-
PR 14-OCT-1999: 99US-0159331-
PR 14-OCT-1999: 99US-0159637-
PR 14-OCT-1999: 99US-0159638-
PR 18-OCT-1999: 99US-0159584-
PR 21-OCT-1999: 99US-0160741-
PR 21-OCT-1999: 99US-0160767-
PR 21-OCT-1999: 99US-0160768-
PR 21-OCT-1999: 99US-0160770-
PR 21-OCT-1999: 99US-0160814-
PR 21-OCT-1999: 99US-0160815-
PR 22-OCT-1999: 99US-0160980-
PR 22-OCT-1999: 99US-0160981-
PR 22-OCT-1999: 99US-0160989-
PR 22-OCT-1999: 99US-0161404-
PR 25-OCT-1999: 99US-0161405-
PR 25-OCT-1999: 99US-0161406-
PR 25-OCT-1999: 99US-0161359-
PR 26-OCT-1999: 99US-0161360-
PR 26-OCT-1999: 99US-0161361-
PR 26-OCT-1999: 99US-0161920-
PR 28-OCT-1999: 99US-0161920-
PR 28-OCT-1999: 99US-0161993-
PR 29-OCT-1999: 99US-0162142-

Query Match

3 6%: Score 99: DB 21: Length 1002:

Best Local Similarity: 19.7%: Pred. No. 1.6: Indels 164: Gaps 20:

Matches 91: Conservative 62: Mismatches 145:

131 KYVEALRYWASTHMSPSWLLKLCDDVPQYFDLITCOLIGRAEVDGASFLKWDGIRTP 190

```

Db      2 KIIIIIIIIIIbhsftvsk-----pitellhallsksftldshspiltskn-1stc 55
Oy      191 FPRSRVLIODL-----VLEGLSHIHRLQNL-----ID1 220
Db      56 fcsvltcdvsilrhvtsldsqhnsqtlssdvahppllqnslaanqisqripqism 115
Oy      221 LVYDNHVAHVASLKVGSE-----LRPLSLH-----TKLOSMSENOTMISLEFHLH 266
Db      116 lyeirhnlnsmvtnngsfpoelssgvlvnlvdlynnmltqdlpvsfntltqlr---hlh 172
Oy      267 GGTSYGFG-----LRVLPESNSDVOLKDLKDLSEANLITANOHSDVICOSEPDSF 315
Db      173 lqnyfsgkipatygtvpyvleylavsgneltg-kippeignlttrellygyvnafehg 231
Oy      316 PSS-GSVSLXVEHRCOLSAIILIDHOYLERPLCALIKOKAPOQVIRAKLSYKPPRL 374
Db      232 ppeigns-----239
Oy      375 FOSYKLCPRKCHLLOEVPHF---GDLILFQD-----GAKTIPDYKLCW 415
Db      240 -elvtidaancqjltgeipeigkikidltfignatgyltqelqllsslkemdl---s 295
Oy      415 TSLYDSKIMTKVQKGRKVAHFKNN--GILPLSECLLIEGTLSEICKL---SKRF 470
Db      296 nmmtgelpstfsgjknltllnlfknklygaip---efl-----gempelerlqikemntf 347
Oy      471 NSVLPFVPSGHED-LELLDLSAPFLIOCTIMHVGTCVTPIC 511
Db      348 tgsipqklgenqrlvildlssnkl-----tqlpnmcc 380

RESULT 15
ID      AAB47250
AA      AAB47250 standard: Protein: 653 AA.
AC      AAB47250:
DI      18-JUL-2001 (first entry)
LE      Human Pp7.
KW      Mitochondrial calcium-binding protein; p95.6/YNS2; calcium:
KM      drug screening; sequence extraction; Pp7; calcium binding motif.
XX      Homo sapiens.
XX      OS
XX      FN      W0200130830-A2.
XX      PD      03-MAY-2001.
XX      FE      26-OCT-2000; 2000KC-US23787.
XX      FP      26-OCT-1999; 59US-0427867.
XX      FR      26-OCT-1999; 59US-0427895.
XX      FR      07-SEP-2000; 2000US-0557253.
XX      FA      (MITO-) MITOKOR.
XX      PI      Murphy AN, Fahy ED, Wiley SE, Clevenger W.
XX      DR      WPI: 2001-308615/32.
XX      DR      N-PSDB: AAC85804.
XX      CC      New Caenorhabditis elegans mitochondrial calcium-binding protein,
PI      p95.6/YNS2, identified by protein motif database searching, useful for
PT      regulating intracellular calcium ions and in drug screening assays.
XX      Example 2; Fig 13; 176pp: English.
XX      PS
XX      CC      This sequence shows a protein containing calmodulin/calcium
XX      CC      binding motif which was identified using the method of the invention.
XX      CC      The method comprises a computer system for extracting information
XX      CC      from biological sequence databases, which receives a motif that defines

```

```

CC      a desired polypeptide sequence. The computer searches the database
CC      based on a query formulated from the motif to obtain a set of
CC      polypeptide sequences which are then filtered and a report of the
CC      filtered set of polypeptide sequence information is then displayed.
CC      The computer system is useful for identifying recombinant C.elegans
CC      p95.6/YNS2 and other mitochondrial proteins involved in binding,
CC      transport or other regulation of intracellular calcium which provides
CC      targets for screening assays that identify agents capable of altering
CC      mitochondrial activity.
XX      -end-
SO      Sequence 653 AA:

```

```

Query Match      3.6% Score 98.5; DB 22; Length 653;
Best Local Similarity 19.3%; Pred.No.0.86;
Matches 82; Conservative 51; Mismatches 118; Indels 173; Gaps 21;

```

```

Oy      120 SKYFNFTED-----IKMVALR---VMASTHSPSWTLKLCQDV-----PMQVFDLTC- 166
Db      63 stfslmtenythlhkeelngslseseqdmrdwdvdsidvpsdysnprlqfplct 119
Oy      157 -----QLGKAEVDGASFLLKWDQIRTF-----PSMRVLI-----QD 200
Db      120 didllleatkeqgllhahvylevfeatkvikmpnfthqtspskevltcgdlhgkld 179
Oy      201 L-----VLEGL-----SHIHRLQNLITDLYDNHVAHVAR-----231
Db      180 lflfyknglpsernpyvfnqdfvdrgrkslellmlcstflvypndhlhrgnhdmm 239
Oy      232 SLKVG---SFLRISLHTR--LQSMNS-----ENOTMISLEFHLHGTSYGRGI 275
Db      240 nlrygtkellhkykhqkrlqileefyawlpjqtivdneliv-----lhqgisetldl 294
Oy      276 RVL-----PESNSDVOLKDLKDLSEANLITANOHSDVICOSEPDSFSSGS 320
Db      295 nllhvrernkmksvllppteirndtdskh-nkygvfnahgrytkngspelen-----347
Oy      321 VSLYEVEHRCOLSAIILIDHOYLE-----RTPLCAL---LKOKAP 357
Db      348 -----llehexeqjldllwsdprgkngcfntcrggacyfqpdtaskll 391
Oy      358 QOYRIAPKLRSY--KPRRLFOSVKLHCPKCHLLOEVPHFEGDLILF-----QNGATKTP 409
Db      392 nkygikmlirshckepg-----yeichdgkvvlftsaasnyeeegsnrga 436
Oy      410 DVKL 413
Db      437 YIKI 440

```

```

Search completed: April 9, 2002, 17:08:16
Job time: 595 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:10:41 ; Search time 64.84 Seconds

(without alignments)
757.005 Million cell updates/sec

Title: US-09-816-248-15

Perfect score: 1775

Sequence: 1 MSWPATNVIYITPLNQLKGG.....LRPGSSVARSRLIASTIS 340

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREML_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1560	93.5	634	4	09NM19
2	1560	93.5	634	4	09NM19
3	348	19.6	384	4	09H662
4	324	18.2	113	11	09CYST7
5	167.5	9.4	555	3	013988
6	153	8.6	463	10	09FNM7
7	144	8.1	497	5	076380
8	130	7.3	251	5	045595
9	111	6.2	759	3	09C2K8
10	109	6.1	587	3	09C3702
11	102.5	5.8	1044	4	09Y4E4
12	103	5.8	4370	4	09H3V5
13	100.5	5.7	215	5	09XWB2
14	98	5.5	1198	10	004538
15	96	5.4	632	5	015742
16	95	5.4	752	5	09GUX3
17	95.5	5.4	474	10	049474
18	94.5	5.3	442	5	09VCU1
19	94.5	5.3	548	5	09NKD4

20	92.5	5.2	1124	1	09HK03	09HK03 thermoplasm
21	92	5.2	607	10	09LS89	09LS89 arabidopsis
22	90.5	5.1	610	1	09HPP8	09HPP8 halobacteri
23	90.5	5.1	1030	10	09C905	09C905 arabidopsis
24	90	5.1	361	2	09C125	09C125 lactococcus
25	90	5.1	937	10	09LNX9	09LNX9 arabidopsis
26	89.5	5.0	528	5	016628	016628 caenorhabdi
27	89.5	5.0	667	10	09SGV4	09SGV4 arabidopsis
28	89	5.0	683	2	084336	084336 treponema p
29	89	5.0	1893	1	09PPA8	09PPA8 uncultured
30	89	5.0	3796	6	09TTK4	09TTK4 bos taurus
31	88.5	5.0	1217	5	09XV62	09XV62 caenorhabdi
32	88	5.0	3788	11	097412	097412 mus musculu
33	88	5.0	3796	6	097699	097699 bos taurus
34	87.5	4.9	497	13	091B98	091B98 potamoxygo
35	87.5	4.9	553	3	09HFK5	09HFK5 neurospora
36	87.5	4.9	564	2	053127	053127 mycobacteri
37	87.5	4.9	1406	13	09W6V5	09W6V5 galus gall
38	87	4.9	387	2	09CM28	09CM28 pasteurella
39	87	4.9	361	10	09MA09	09MA09 phytomyza i
40	87	4.9	484	8	09G4L4	09G4L4 oryza sativ
41	87	4.9	723	10	09SNK7	09SNK7 arabidopsis
42	85.5	4.9	318	10	09S137	09S137 arabidopsis
43	85.5	4.9	1002	10	09M2Z1	09M2Z1 arabidopsis
44	85	4.8	175	3	09C2X7	09C2X7 candida alb
45	86	4.8	504	8	09TB28	09TB28 greya punct

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	634 AA.
09NM19	09NM19			
ID	09NM19			
DI	01-OCT-2000 (TRENBERG)	15, Created		
DI	01-OCT-2000 (TRENBERG)	15, Last sequence update		
DI	01-OCT-2000 (TRENBERG)	15, Last annotation update		
DE	CDNA FLJ10368 FIS, CLONE NTZRM2001544.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
OX	NCRL_TaxID=9605;			
RN	[1]			
RF	SEQUENCE FROM N.A.			
RA	Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,			
RA	Wakatsuna M., Hosofri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,			
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,			
RA	Matanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,			
RA	Tamamoto J., Nakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,			
RA	Ninomiya K., Iwayanagi T.,			
RI	"NEBO human cDNA sequencing project."			
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AK001230; BAA91568.1; -			
SO	SEQUENCE	634 AA: 71425 MW: C0A9C40823F9540C CRC64:		

Query Match	93.5%	Score 1660:	DB 4:	Length 634:
Best local Similarity	98.8%	Pred. 1e-138:		
Matches	317:	Conservative	2:	Mismatches
			2:	Indels
				Gaps
				0:
				C:
09	1	MSWPATNVIYITPLNQLKGGIYVNVGVVFFKPPYLSKTDICSVYIVDQNVNLTCL	60	
DB	1	MSWPATNVIYITPLNQLKGGIYVNVGVVFFKPPYLSKTDICSVYIVDQNVNLTCL	60	
09	61	LFSGNVYALPIYKNGDVFHRLKIOVKKETOGITSSGFASLTFEGTIGAPIIRIS	120	
DB	61	LFSGNVYALPIYKNGDVFHRLKIOVKKETOGITSSGFASLTFEGTIGAPIIRIS	120	
09	121	KYFNFTEDHKRWKVALRWASTHMSPSWTLKLCDOVPMOYFDLTCOLGKAEVDCASFL	180	
DB	121	KYFNFTEDHKRWKVALRWASTHMSPSWTLKLCDOVPMOYFDLTCOLGKAEVDCASFL	180	

```

OY 181 LKWDGRTPEPSRNVLLDQVLVEGDSLHRLQNLITDILVYDNRHVAFSLKVGSEFLR 240
Db 181 LKWDGRTPEPSRNVLLDQVLVEGDSLHRLQNLITDILVYDNRHVAFSLKVGSEFLR 240
OY 241 IYSLHTKLSQMSXSENOTMLSLFPHLHGISTYGRGIPVLPESNSDVPDLKADLESANLTAN 300
Db 241 IYSLHTKLSQMSXSENOTMLSLFPHLHGISTYGRGIPVLPESNSDVPDLKADLESANLTAN 300
OY 301 QHSDVICOSEFDDSPFGVSL 321
Db 301 QHSDVICOSEFDDSPFGSV 321
RESULT 2
O9NMJX5 PRELIMINARY: PRI: 634 AA.
AC O9NMJX5:
DI 01-OCT-2000 (TfEMBLrel. 15, Created)
DI 01-OCT-2000 (TfEMBLrel. 15, Last sequence update)
DI 01-JUN-2001 (TfEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 71.4 KDA PROTEIN (UNKNOWN) (PROTEIN FOR MCC:1028.).
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Cranialia: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
OX NCBI_TaxID:9606:
PN [1]
PP SEQUENCE FROM N.A.
PC TISSUE=PLACENTA:
PA Isogal T., Ota T., Hayashi K., Sugiyama T., Otsuki I., Suzuki Y.,
PA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
PA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakematsu A.,
PA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.:
PI "NEDO human cDNA sequencing project."
PI Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
PN [2]
PP SEQUENCE FROM N.A.
PA Isogal T., Ota T., Hayashi K., Sugiyama T., Otsuki I., Suzuki Y.,
PA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
PA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakematsu A.,
PA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.:
PI "NEDO human cDNA sequencing project."
PI Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
PN [3]
PP SEQUENCE FROM N.A.
PC TISSUE=SKIN, AND MELANOMA:
PA Strassberg R.:
PI Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK001935; PARS1988.1;
DR EMBL: AK022580; BAB14110.1;
DR EMBL: BC002923; AAH02923.1;
DR SEQUENCE 534 AA: 71441 MM: 123A12CABE708C93 CPG54:
Query Match 93.5%: Score 1660: DB 4: Length 634:
Best Local Similarity 98.8%: Pred. No. Jc:139:
Matches 317: Conservative 2: Mismatches 2: Indels 0: Gaps
OY 1 MSLVPAINTIYITPLNOLKSGTIVNVGVGVFFRPYLSKSDICDYSVIIVDQTNKICL 60
Db 1 MSLVPAINTIYITPLNOLKSGTIVNVGVGVFFRPYLSKSDICDYSVIIVDQTNKICL 60
OY 61 LFGSNVYALPIYIKRNGDIVFHFHLKIOVYKKEIQJITSSGFASLTFFEGTLGAPITPES 120
Db 61 LFGSNVYALPIYIKRNGDIVFHFHLKIOVYKKEIQJITSSGFASLTFFEGTLGAPITPES 120
OY 121 KYENFTTEDHKRNEALRYWASTHMSPSWTLKLCQDVQPMQYEDTLCOLLGKAEPVDSFL 180
Db 121 KYENFTTEDHKRNEALRYWASTHMSPSWTLKLCQDVQPMQYEDTLCOLLGKAEPVDSFL 180
OY 181 LKWDGRTPEPSRNVLLDQVLVEGDSLHRLQNLITDILVYDNRHVAFSLKVGSEFLR 240
Db 181 LKWDGRTPEPSRNVLLDQVLVEGDSLHRLQNLITDILVYDNRHVAFSLKVGSEFLR 240

```

Q7	241	IYSLKTIKIQSMNSEKQITLSEFLHIGCTSYGRGIRVLPENSNDVDLKKLECANITAN	300
Db	241	IYSLHITKIQSMNSEKQITLSEFLHIGCTSYGRGIRVLPENSNDVDLKKLECANITAN	300
Q7	301	QHSVDYICQSEPPDSFPNGSVL	321
Db	301	QHSVDYICQSEPPDSFPSSGSV	321

```

RESULT      3
C9H652
ID      O9H652      PRELIMINARY;      PRI:      384 AA.
AC      O9H652
DT      01-MAR-2001 (TReMBLrel 15, Created)
DT      01-MAR-2001 (TReMBLrel 15, Last sequence update)
DT      01-MAR-2001 (TReMBLrel 15, Last annotation update)
DE      CDNA: FLJ22581 FIS, CLONE HS102573.
DS      Homo sapiens (Human).
CC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX      NCBI_taxid=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=HUMAN SMALL INTESTINE;
RA      Watanabe K., Kunagaki A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA      Suzuki Y., Ohayashi M., Nishi T., Shibahara T., Tanaka T.,
RA      Nakamura Y., Isono T., Sugano S.;
RT      "RED human cDNA sequencing project.";
RL      Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AK026224; BAF15404.1;
SO      SEQUENCE      384 AA: 13161 MW: 60C968D595CECAC6 CRC64:

```

```

Query March 19.6% Score 348: DB 4: Length 384:
Best Local Similarity 94.4%: Pred. No. 1, le-22:
Matches 67: Conservative 2: Mismatches 2: Indels 0: Gaps 0:

Q: 251 MNSNTOIMSLFEHFKGCTSYGSGTGVLPESNSDDVQLKKDLESANLTAQHSQVTCQSE 310
      |||||
Db 1 MNSNTOIMSLFEHFKGCTSYGSGTGVLPESNSDDVQLKKDLESANLTAQHSQVTCQSE 60
      |||||

Q: 311 FDDSFPSGSEV 321
      |||||
Db 61 FDDSFPSGSEV 71

```

RESULT	4	
09CYST7		
ID	09CYST7	PRELIMINARY:
CD	09CYST7	PRT: 113 AA.
DT	01-JUN-2001 (TREMBlrel. 17, Created)	
DI	01-JUN-2001 (TREMBlrel. 17, Last sequence update)	
D1	01-JUN-2001 (TREMBlrel. 17, Last annotation update)	
DE	2810458H16RIK PROTEIN.	
GN	2810458H16RIK.	
OS	Mus musculus (Mouse).	
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
CX	NCBI_TaxID=10090:	
FX	113	
RP	SEQUENCE FROM N.A.	
RF	STFAIR-CSTAL/63: TISSUE=EMBRYO;	
RC	MEELINE-J10855690: PubMed=11217851:	
RX	Kawai J., Shingaya A., Shibata K., Yoshino M., Itoh M., Ishii Y.,	
PA	Aizawa K., Hara A., Fukunishi Y., Komio H., Adachi J., Fukuda S.,	
PA	Aizawa K., Iizawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,	
PA	Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,	
PA	Kodjima K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,	
PA	Fiedlerchann M., Gaasterland T., Gissi C., King B., Kochiya H.,	
PA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,	
PA	Schirral L.M., Steubli F., Suzuki R., Tomita M., Wagner L., Washio T.,	
PA	Sekal K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,	

	Matches	503:	Conservative	1:	Mismatches	3:	Indels	n:	Gaps
OY	1	MSLVPAIRNLYIPLNOLKGGIIVNNVGVVKFEKKPEPLYSKGIDGCVSVTIVDOINWKL	60						
DH	1	MSLVPAIRNLYIPLNOLKGGIIVNNVGVVKFEKKPEPLYSKGIDGCVSVTIVDOINWKL	60						
OY	61	LFGNGEALPIIKRNDIYFHRHLKIQVYKRKTGGITSSGFASLIFEGTIGAPLIIFPTSS	120						
DH	61	LFGNGEALPIIKRNDIYFHRHLKIQVYKRKTGGITSSGFASLIFEGTIGAPLIIFPTSS	120						
OY	121	KYNFTTEDHKVEALRVMASTHMSPSWTLKLCDVPOMGYFDLICOLLGAEYDGAFL	180						
DH	121	KYNFTTEDHKVEALRVMASTHMSPSWTLKLCDVPOMGYFDLICOLLGAEYDGAFL	180						
OY	181	LKWMDGTFRPFEPKARVLDDVLEEDLSIHRLNLVIDLVDRHHVAASLKVSSELR	240						
DH	181	LKWMDGTFRPFEPKARVLDDVLEEDLSIHRLNLVIDLVDRHHVAASLKVSSELR	240						
OY	241	IYSLHTKLOSNGENOTMLSEFHLLHGSGSYRGIRVLPESNSDVDLCKDKLEFAN	300						
DH	241	IYSLHTKLOSNGENOTMLSEFHLLHGSGSYRGIRVLPESNSDVDLCKDKLEFAN	300						
OY	301	QHSVDICQSEPDSPFSQSGSVSLYEVEPCQLSATILLDHQYLEFTPLCALIKKRAPOOY	360						
DH	301	QHSVDICQSEPDSPFSQSGSVSLYEVEPCQLSATILLDHQYLEFTPLCALIKKRAPOOY	360						
OY	361	RIPAKLSYPRPRLFOSVKRIHCPCCHLOEVPHEGDIIIFDGSAKTPDYKLOHI	420						
DH	361	RIPAKLSYPRPRLFOSVKRIHCPCCHLOEVPHEGDIIIFDGSAKTPDYKLOHI	420						
OY	421	SKMTTKRKGRKYAAHVFKVNRNGLILPLSNECLLIEGGLSEICKLSKNFSVPIPVPSGH	480						
DH	421	SKMTTKRKGRKYAAHVFKVNRNGLILPLSNECLLIEGGLSEICKLSKNFSVPIPVPSGH	480						
OY	481	EDELLELDLSAFLTGCTIHHYGTCYT	507						
DH	481	EDELLELDLSAFLTGCTIHHYGTCYT	507						
RESULT 2									
ID	09NM19	PREFINITARY:	PRT:	634 AA.					
AC	09NM19:								
DT	01-OCT-2000 (IREMBrel_15, Created)								
DT	01-OCT-2000 (IREMBrel_15, Last sequence update)								
DT	01-OCT-2000 (IREMBrel_15, Last annotation update)								
DE	CDNA FLJ10368 fts, CLOCN NTZEM2001544.								
DC	Homo sapiens (Human)								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
FN	NCBL TaxID=9605;								
PN	[1]								
PR	SEQUENCE FROM N.A.								
PA	Isoagi T., Ota T., Hayashi K., Sugiyama T., Otsuki I., Suzuki Y.,								
PA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,								
PA	Takeshima M., Hosohiti T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,								
PA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,								
PA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Satoh K.,								
PA	Yamamoto J., Wakatsutsu A., Nakamura Y., Nagahari K., Masuno Y.,								
PA	Nimomiya K., Iwayanagi T.								
PT	NCBI human cDNA sequencing project."								
RL	Submitted (Feb-2000) to the EMBL/Genbank/DBJ databases.								
EW	EMBL: AK001230; BAA31568.1;								
SQ	SEQUENCE 534 AA: 71425 MW: CDA9C40B23F9540C CRC64:								

```

Query Match 95.2% Score 2528 DB 4 Length 534:
Fast Local Similarity: 52.0% P-Val 5e-214
Matches 502; Conservative 1; Mismatches 4; Indels 0; Gaps 0
0: 1 MSLVPAATNYYITFLPOLKGT IANNVGVKFFEEFPLSKIDYCSVTIYDQIHNFIL 60
1 MSLVPAATNYYITFLNOLKGT IANNVGVKFFEEFPLSKIDYCSVTIYDQIHNFIL 60

```

```

07 61 LFSGNVALLPIIYKNGDIYVFRHRLKIQYKKETOGITSSGFASLTFEGTIGAPIIPRTSS 120
Db 61 LFSGNVEALPIIYKNGDIYVFRHRLKIQYKKETOGITSSGFASLTFEGTIGAPIIPRTSS 120
07 121 KYFNFTTEDHKMVEALRYWASTHMSPSVTLTKLCDVOPMOYFDTLCOLLGKAEVDGASFL 180
Db 121 KYFNFTTEDHKMVEALRYWASTHMSPSVTLTKLCDVOPMOYFDTLCOLLGKAEVDGASFL 180
07 181 LKXWVDGRTPEPSKRVLIODLVLEGDSLHHRLOMLTIDLIVYDNHVAVASLKVGSFLR 240
Db 181 LKXWVDGRTPEPSKRVLIODLVLEGDSLHHRLOMLTIDLIVYDNHVAVASLKVGSFLR 240
07 241 IYSLHTKLIQSNHSEKOTLSEFLHNGSTVGRGIRVLPESNSVDOLKRLSESANLTAN 300
Db 241 IYSLHTKLIQSNHSEKOTLSEFLHNGSTVGRGIRVLPESNSVDOLKRLSESANLTAN 300
07 301 QHSVDYICSEPDSPDFSPSSGSVSEVEVERCOOLSATILTDHOYLERTPICALIKOKAPOY 360
Db 301 QHSVDYICSEPDSPDFSPSSGSVSEVEVERCOOLSATILTDHOYLERTPICALIKOKAPOY 360
07 361 RIRAKLRSTKPRPLFOSKYLHPKCHLLOEVNHEGDDLIIFODGATKTPDVOKLONTSLYD 420
Db 361 RIRAKLRSTKPRPLFOSKYLHPKCHLLOEVNHEGDDLIIFODGATKTPDVOKLONTSLYD 420
07 421 SKIWTIKNKGKRVAVHFEKRNJILPLSNECLLIEGGLSEICLSKFNFSVIVPVSQH 480
Db 421 SKIWTIKNKGKRVAVHFEKRNJILPLSNECLLIEGGLSEICLSKFNFSVIVPVSQH 480
07 481 EDLELLDISAPFLIOGTIHNYCTGYCT 507
Db 481 EDLELLDISAPFLIOGTIHNYCTGYCT 507

RESULT 3
09H562 ID 09H562 PRELIMINARY: PRT: 384 AA.
AC 09H562:
DT 01-MAR-2001 (TREMBlrel. 15, Created)
DT 01-MAR-2001 (TREMBlrel. 15, last sequence update)
DT 01-MAR-2001 (TREMBlrel. 15, last annotation update)
CDNA: FLJ22581 FIS, CLONE HS102573.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9605;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HUMAN SMALL INTESTINE;
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohtsuyashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isono I., Sugano S.;
RI "RED human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBU databases.
EMBL: AK029234. BAB15404.1.
SQ
SEQUENCE 384 AA: 43161 PIR: 60CG68059SCCEAC6 CR664:

```

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:10:44 ; Search time 64.84 Seconds
(without alignments)

1168.554 Million cell updates/sec

Title: US-09-816-248-17

Sequence: 1 MSLVPAINVITPLNOLKGG.....HHYGTGCTPPIVCYDLYT 518

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organelle:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2637	96.5	634	4 Q9NM19	Q9NM19 homo sapien
2	2628	96.2	634	4 Q9NM19	Q9NM19 homo sapien
3	1371	48.4	384	4 Q9NM62	Q9NM62 homo sapien
4	976	35.7	317	4 Q9NM18	Q9NM18 homo sapien
5	324	11.9	113	11 Q9C1S7	Q9C1S7 mus muscu
6	201.5	7.4	555	3 Q13988	Q13988 schizosacch
7	153.5	5.6	463	10 Q9FNH7	Q9FNH7 arabidopsi
8	144.5	5.3	497	5 Q76380	Q76380 oxytricha
9	130	4.8	587	5 Q93702	Q93702 caenorhabd
10	117.5	4.3	251	5 Q45595	Q45595 caenorhabd
11	115	4.2	115	12 Q9QMH2	Q9QMH2 parvo-like
12	111	4.1	759	3 Q9C1K8	Q9C1K8 neurospira
13	110.5	4.0	483	10 Q9XE80	Q9XE80 sorghum bic
14	110.5	4.0	1044	4 Q9Y4E4	Q9Y4E4 homo sapien
15	108	4.0	1217	5 Q9XV62	Q9XV62 caenorhabd
16	106	3.9	2391	5 Q27732	Q27732 plasmodium
17	105.5	3.9	548	5 Q9AKD4	Q9AKD4 drosophila
18	105	3.8	655	4 Q12890	Q12890 homo sapien
19	104	3.8	1112	12 Q9YUR6	Q9YUR6 turkey aden

20	103.5	3.8	493	2 Q9JMH8	Q9JMH8 wolbachia s
21	103.5	3.8	892	10 Q9XEJ5	Q9XEJ5 zea mays (m
22	103	3.8	896	2 Q9CGJ2	Q9CGJ2 lactococcus
23	103	3.8	1434	5 Q62239	Q62239 caenorhabd
24	102.5	3.8	520	13 Q91R52	Q91R52 xenopus lae
25	102.5	3.8	528	5 Q16528	Q16528 caenorhabd
26	102.5	3.8	698	10 Q9SLM4	Q9SLM4 arabidopsi
27	102	3.7	4370	4 Q9H3V5	Q9H3V5 homo sapien
28	101.5	3.7	737	10 Q9SJ13	Q9SJ13 arabidopsi
29	101.5	3.7	1038	2 Q99RD2	Q99RD2 staphylococ
30	101	3.7	442	5 Q9YUJ1	Q9YUJ1 drosophila
31	101	3.7	773	10 Q9SHP2	Q9SHP2 arabidopsi
32	100.5	3.7	1438	3 Q06681	Q06681 saccharomyc
33	100	3.7	1110	10 Q9F177	Q9F177 arabidopsi
34	99	3.6	684	2 Q9S3C2	Q9S3C2 porphyromon
35	99	3.6	712	8 Q19880	Q19880 cyanidium c
36	99	3.6	1002	10 Q9K2Z1	Q9K2Z1 arabidopsi
37	99	3.6	1428	5 Q77023	Q77023 drosophila
38	98.5	3.6	735	5 Q9N8A6	Q9N8A6 trypanosoma
39	98.5	3.6	1003	5 Q9TXJ6	Q9TXJ6 leishmania
40	98.5	3.6	1619	5 Q77382	Q77382 plasmodium
41	98	3.6	909	5 Q17429	Q17429 caenorhabd
42	98	3.6	2747	5 Q9BUX9	Q9BUX9 plasmodium
43	97.5	3.6	755	12 Q56271	Q56271 human herpes
44	97	3.6	372	2 Q24967	Q24967 helicobacte
45	97	3.6	1123	10 Q9ESL9	Q9ESL9 arabidopsi

ALIGNMENTS

RESULT 1
Q9NM19
AC Q9NM19: PRELIMINARY: PRT: 634 AA.
DI 01-OCT-2000 (IREMBLrel, 15, last sequence update)
DI 01-OCT-2000 (IREMBLrel, 17, last annotation update)
DI 01-JUN-2001 (IREMBLrel, 17, last annotation update)
DE HYPOTHETICAL 71.4 KDA PROTEIN (UNKNOWN) (PROTEIN FOR MGC:10280).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuno Y., Sasaki N.,
RA *Redo human cDNA sequencing project.*
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuno Y., Sasaki N.,
RA *Redo human cDNA sequencing project.*
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Strausberg R., AND MELANOMA:
RA Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK001935; BAA91988.1;
DR EMBL: AK022580; BAB14110.1;
DR EMBL: BC002923; AAH02923.1;
SO SEQUENCE 634 AA: 71441 MW: 123412CABE708C91 CRC64:

Query Match 96.5% Score 2637; DB 4; Length 634;
Best Local Similarity 99.2% Pred. No. 1,2e-214;


```

QY 181 LKWDGTRIFPSPRWVLIODLVLEGLDLSHRLONLTIIDLVDNHHVARSLEKVGSLR 240
    |||||||
DB 181 LKWDGTRIFPSPRWVLIODLVLEGLDLSHRLONLTIIDLVDNHHVARSLEKVGSLR 240
QY 241 IYSLHTRKLOSMNSNOTMLSEFLHMGTSYGRGIRVLPESNSDVDC LKQDESANLTAN 300
    |||||||
DB 241 IYSLHTRKLOSMNSNOTMLSEFLHMGTSYGRGIRVLPESNSDVDC LKQDESANLTAN 300
QY 301 QHSDVICOSEPDSEFPSSGVSLEYVERCOOLSAITILDHOYLERTPLCALLKOKAPOY 360
    |||||||
DB 301 QHSDVICOSEPDSEFPSSGVSLEYVERCOOLSAITILDHOYLERTPLCALLKOKAPOY 360
QY 361 RIRAKLESTYKPRPLFOSVKLHCPKCHLLOEVPHEGDLDIIFOGATKTPVVKLONTSLYD 420
    |||||||
DB 361 RIRAKLESTYKPRPLFOSVKLHCPKCHLLOEVPHEGDLDIIFOGATKTPVVKLONTSLYD 420
QY 421 SKIWTIRKQKPKVAVHFKVKNNGILPLSNECLLIEGGLSEICKLSKFNFSVIVPVRSGH 480
    |||||||
DB 421 SKIWTIRKQKPKVAVHFKVKNNGILPLSNECLLIEGGLSEICKLSKFNFSVIVPVRSGH 480
QY 481 EDLELLDLSAPFLIOGTIHHYGCQCCSLRSIONLNSLVDTSMIPSSVAEALGIVPLOY 540
    |||||||
DB 481 EDLELLDLSAPFLIOGTIHHYGCQCCSLRSIONLNSLVDTSMIPSSVAEALGIVPLOY 540
QY 541 VFVMTFTLDDGTGVLAVYLMDSKDFFOIPASEVLMDDLOKSVDMIMNFCPPGIKIDAY 600
    |||||||
DB 541 VFVMTFTLDDGTGVLAVYLMDSKDFFOIPASEVLMDDLOKSVDMIMNFCPPGIKIDAY 600
QY 601 FWLECFIKSYNVTNGTDNOICYOIFDTTVAEDVI 634
    |||||||
DB 601 FWLECFIKSYNVTNGTDNOICYOIFDTTVAEDVI 634

RESULT 2
Q9N0X5 PRELIMINARY: PRT: 634 AA.
AC 09N0X5:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOHETICAL 71.4 KDA PROTEIN (UNKNOWN) (PROTEIN FOR MGC:10280).
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
OX NCBI_TaxID=9606:
RN 11
FP SEQUENCE FROM N.A.
PC TISSUE=PLACENTA:
PA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki I., Suzuki Y.,
PA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
PA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
PA Nakamura Y., Nagahari K., Masuno Y., Sasaki N.:
PA *NEDO human cDNA sequencing project.*
PL Submitted (Feb-2000) to the EMBL/GenBank/DBJ databases.
RN 12
FP SEQUENCE FROM N.A.
PC TISSUE=SKIN, AND MELANOMA:
PA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki I., Suzuki Y.,
PA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
PA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
PA Nakamura Y., Nagahari K., Masuno Y., Sasaki N.:
PA *NEDO human cDNA sequencing project.*
PL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN 13
FP SEQUENCE FROM N.A.
PC TISSUE=SKIN, AND MELANOMA:
PA Strausberg R.:
PA Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AK001935; BAA91988.1:
DR EMBL: AK022580; BAB14110.1:
DR EMBL: BC002923; AAH02923.1:
SQ SEQUENCE 634 AA: 71441 MW: 123A12CABE708C91 CRC6:

```

```

Query Match 99.8%: Score 3317: DB 4: Length 634:
Best Local Similarity 99.8%: Pred. No. 4, 2e-265:
Matches 633: Conservative 0: Mismatches 1: Indels 0: Gaps 0:
QY 1 MSLVPATNITITPLNOLKGGTIVNVGVKFFKPPPLSKGTGYCSWVITVDOTNKLTC 60
    |||||||
DB 1 MSLVPATNITITPLNOLKGGTIVNVGVKFFKPPPLSKGTGYCSWVITVDOTNKLTC 60
QY 61 LFSGNTEALPIIKNGDIVRFBKLTQVTKKETOGITSSGFASLFEGLTGAIPRTSS 120
    |||||||
DB 61 LFSGNTEALPIIKNGDIVRFBKLTQVTKKETOGITSSGFASLFEGLTGAIPRTSS 120
QY 121 KYNFTTEKMKVAVLRVASTHNSPSWTLKLCVOPQOYFDLCCOLLGKAEVGSASF 180
    |||||||
DB 121 KYNFTTEKMKVAVLRVASTHNSPSWTLKLCVOPQOYFDLCCOLLGKAEVGSASF 180
QY 181 LKWDGTRIFPSPRWVLIODLVLEGLDLSHRLONLTIIDLVDNHHVARSLEKVGSLR 240
    |||||||
DB 181 LKWDGTRIFPSPRWVLIODLVLEGLDLSHRLONLTIIDLVDNHHVARSLEKVGSLR 240
QY 241 IYSLHTRKLOSMNSNOTMLSEFLHMGTSYGRGIRVLPESNSDVDC LKQDESANLTAN 300
    |||||||
DB 241 IYSLHTRKLOSMNSNOTMLSEFLHMGTSYGRGIRVLPESNSDVDC LKQDESANLTAN 300
QY 301 QHSDVICOSEPDSEFPSSGVSLEYVERCOOLSAITILDHOYLERTPLCALLKOKAPOY 360
    |||||||
DB 301 QHSDVICOSEPDSEFPSSGVSLEYVERCOOLSAITILDHOYLERTPLCALLKOKAPOY 360
QY 361 RIRAKLESTYKPRPLFOSVKLHCPKCHLLOEVPHEGDLDIIFOGATKTPVVKLONTSLYD 420
    |||||||
DB 361 RIRAKLESTYKPRPLFOSVKLHCPKCHLLOEVPHEGDLDIIFOGATKTPVVKLONTSLYD 420
QY 421 SKIWTIRKQKPKVAVHFKVKNNGILPLSNECLLIEGGLSEICKLSKFNFSVIVPVRSGH 480
    |||||||
DB 421 SKIWTIRKQKPKVAVHFKVKNNGILPLSNECLLIEGGLSEICKLSKFNFSVIVPVRSGH 480
QY 481 EDLELLDLSAPFLIOGTIHHYGCQCCSLRSIONLNSLVDTSMIPSSVAEALGIVPLOY 540
    |||||||
DB 481 EDLELLDLSAPFLIOGTIHHYGCQCCSLRSIONLNSLVDTSMIPSSVAEALGIVPLOY 540
QY 541 VFVMTFTLDDGTGVLAVYLMDSKDFFOIPASEVLMDDLOKSVDMIMNFCPPGIKIDAY 600
    |||||||
DB 541 VFVMTFTLDDGTGVLAVYLMDSKDFFOIPASEVLMDDLOKSVDMIMNFCPPGIKIDAY 600
QY 601 FWLECFIKSYNVTNGTDNOICYOIFDTTVAEDVI 634
    |||||||
DB 601 FWLECFIKSYNVTNGTDNOICYOIFDTTVAEDVI 634

RESULT 3
Q9H662 PRELIMINARY: PRT: 384 AA.
AC 09H662:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 15, Last annotation update)
DE CDNA: FLJ22581 FIS, CLONE HS102573.
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
OX NCBI_TaxID=9606:
RN 11
FP SEQUENCE FROM N.A.
PC TISSUE=HUMAN SMALL INTESTINE:
PA Watanabe K., Kunagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
PA Suzuki Y., Chayashi M., Nishi T., Shibahara T., Tanaka T.,
PA Nakamura Y., Isegai T., Sugano S.:
PA *NEDO human cDNA sequencing project.*
PL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AK026234; BAB15404.1:
DR EMBL: AK026234; BAB15404.1:
SQ SEQUENCE 384 AA: 43161 MW: 6CC968D595CECAC6 CRC64:

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2002, 17:06:41 ; Search time 64.84 Seconds

(without alignments)
1430.238 Million cell updates/sec

Title: US-09-816-248-13

Sequence: 1 MSLVPAIRNYITPLNOLKGG.....GIDNOICVQIEDTIVAEV 634

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL.17:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.virus:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3324	100.0	634	4	Q9NM19
2	3317	99.8	634	4	Q9NM19
3	2001	50.2	384	4	Q9NM19
4	1655	49.8	317	4	Q9NM19
5	716	21.5	135	4	Q9NM19
6	324	9.7	113	11	Q9CY57
7	199.5	6.0	463	10	Q9FNH7
8	153.5	4.3	497	5	Q6380
9	144.5	4.3	497	5	Q6380
10	130.5	3.9	587	5	Q63702
11	123	3.7	115	12	Q9GMH2
12	122	3.7	114	10	Q40392
13	118	3.5	1217	5	Q9XV62
14	117.5	3.5	251	5	Q45595
15	116	3.5	1044	4	Q9V4E4
16	114.5	3.4	745	3	Q9P7X5
17	114.5	3.4	1110	10	Q9T177
18	114	3.4	899	10	Q9SG75
19	114	3.4	917	10	Q9CAF6

20	114	3.4	1348	5	Q9VAD1
21	114	3.4	1475	10	Q9SXQ4
22	114	3.4	2391	5	Q27732
23	112.5	3.4	1112	10	Q9S752
24	112	3.4	907	10	Q9FM49
25	112	3.4	1421	10	Q9SXQ5
26	112	3.4	2528	5	Q19317
27	111.5	3.4	483	10	Q9XE80
28	111	3.3	759	3	Q9C2K8
29	111	3.3	1447	10	Q9SXQ3
30	110	3.3	896	2	Q9CGJ2
31	109.5	3.3	1456	10	Q9Z794
32	109	3.3	1534	10	Q9FZK7
33	108.5	3.3	892	10	Q9XEJ5
34	108	3.2	927	2	Q31052
35	107.5	3.2	980	10	Q9ZU10
36	107.5	3.2	1784	10	Q9C6R1
37	107	3.2	501	1	Q26978
38	106	3.2	773	10	Q9SHP2
39	106	3.2	1123	10	Q9SSL9
40	106	3.2	1265	5	Q9VBR8
41	105.5	3.2	548	5	Q9NRD4
42	105.5	3.2	1038	2	Q99RD2
43	105.5	3.2	6420	2	P95814
44	105	3.2	695	5	Q9VSN5
45	105	3.2	1112	12	Q9YUR6

ALIGNMENTS

RESULT	ID	Q9NM19	PRELIMINARY:	PRI:	634 AA.
AC	Q9NM19				
DT	01-OCT-2000	(TRENBLREL. 15, Created)			
DT	01-OCT-2000	(TRENBLREL. 15, Last sequence update)			
DT	01-OCT-2000	(TRENBLREL. 15, Last annotation update)			
DE	CDNA FL010368 FIS, CLONE NTZRM201544.				
OS	Homo sapiens (Human).				
OC	Eukaryota: Chordata: Vertebrata: Euteleostomi:				
OC	Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Isoqal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,				
RA	Mishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,				
RA	Wagatsuna M., Hosokawa T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,				
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,				
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,				
RA	Yamanoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,				
RI	Ninomiya K., Iwayanagi T.,				
RI	"NEBO human cDNA sequencing project."				
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AK001230; BAA91568.1;				
SO	SEQUENCE 634 AA: 71425 MW: C0A9C4082F9540C CRC64:				

Query Match	Score	3324	DB 4:	Length	634:
Best local Similarity	100.0%	Pred. No.	1.1e-265:		
Matches	634:	Conservative	0:	Mismatches	0:
				Indels	0:
				Gaps	0:
QY	1	MSLVPAIRNYITPLNOLKGGIIVNYGVKFRPKPLSGDYGCVIIVDQINYLKCL	60		
DB	1	MSLVPAIRNYITPLNOLKGGIIVNYGVKFRPKPLSGDYGCVIIVDQINYLKCL	60		
QY	51	LSFGNEALPIIYKNDIVFRHLKIQVYKKETOGITSSGFASLTFFGLGAPILPRTS	120		
DB	51	LSFGNEALPIIYKNDIVFRHLKIQVYKKETOGITSSGFASLTFFGLGAPILPRTS	120		
QY	121	KTFNFTEDHKWEALRWASTHMSPSWLLKLCVQVPOHFDLTCLLGKAEVDGASFL	180		
DB	121	KTFNFTEDHKWEALRWASTHMSPSWLLKLCVQVPOHFDLTCLLGKAEVDGASFL	180		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 12, 2002, 08:33:16 ; Search time 22.44 Seconds

(without alignments)
710,503 Million cell updates/sec

Title: US-09-816-248-5
Perfect score: 567
Sequence: 1 MSLEPAINIVITPLNOLKGG.....KKETOGITSSGFASLTFEET 109

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_17: *
2: sp_archaea: *
3: sp_bacteria: *
4: sp_fungi: *
5: sp_human: *
6: sp_invertebrate: *
7: sp_mammal: *
8: sp_mhc: *
9: sp_organelle: *
10: sp_phage: *
11: sp_plant: *
12: sp_rodent: *
13: sp_virus: *
14: sp_unclassified: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	567	100.0	634	4 09NMU9	09NMU9 homo sapien
2	567	100.0	634	4 09NMU5	09NMU5 homo sapien
3	324	57.1	113	11 09C1S7	09C1S7 mus musculu
4	114	20.1	555	3 013988	013988 schizosacch
5	106	18.7	427	5 076380	076380 oxytricha t
6	85	15.0	318	10 09S137	09S137 arabidopsis
7	77.5	13.7	473	5 09Y119	09Y119 cytoplasmic
8	75	13.2	759	3 09C2K8	09C2K8 neurospera
9	74.5	13.1	383	11 09COE7	09COE7 mus musculu
10	74	13.1	198	4 09S330	09S330 mus sapien
11	74	13.1	228	4 09BON7	09BON7 homo sapien
12	74	13.1	345	4 09PIL3	09PIL3 homo sapien
13	74	13.1	383	4 09Y282	09Y282 homo sapien
14	72.5	12.8	299	5 09NF41	09NF41 drosophila
15	72.5	12.8	937	10 09LNX9	09LNX9 arabidopsis
16	71.5	12.6	1054	2 09KRF1	09KRF1 bacillus ha
17	70.5	12.5	247	2 09ZFO1	09ZFO1 buchiera ap
18	70.5	12.4	287	2 046329	046329 cowdria rum
19	70.5	12.4	287	2 046331	046331 cowdria rum

20	70.5	12.4	287	2 09R425	09R425 cowdria rum
21	69.5	12.3	313	2 09AKM8	09AKM8 rickettsia
22	69.5	12.3	313	2 09AKH3	09AKH3 rickettsia
23	69.5	12.3	383	2 P74756	P74756 synchocyst
24	69.5	12.3	467	2 034798	034798 bacillus su
25	69.5	12.3	886	10 09Z007	09Z007 arabidopsis
26	68.5	12.1	812	2 09RT53	09RT53 delnococtus
27	68.5	12.1	1124	1 09HR03	09HR03 thermoplasm
28	68.5	12.0	792	3 09HC45	09HC45 kluveromyc
29	67.5	11.9	306	2 09KRF9	09KRF9 bacillus ha
30	67.5	11.9	877	12 099BV2	099BV2 baboon cito
31	67	11.8	480	5 017840	017840 caenorhabdi
32	67	11.8	497	2 09K3R9	09K3R9 streptomyce
33	67	11.8	518	5 P91345	P91345 caenorhabdi
34	67	11.8	1175	2 047955	047955 haemophilus
35	66.5	11.7	490	5 022200	022200 caenorhabdi
36	66.5	11.7	664	3 09URD2	09URD2 schizosacch
37	66.5	11.7	801	3 09P3N3	09P3N3 schizosacch
38	66	11.6	279	13 09PUS5	09PUS5 brachydanio
39	66	11.6	576	10 09LIR8	09LIR8 plasmodium
40	65	11.5	2867	5 09N2M3	09N2M3 plasmodium
41	65.5	11.5	284	2 09AF41	09AF41 cowdria rum
42	65.5	11.6	313	2 09AKC5	09AKC5 rickettsia
43	65.5	11.5	598	12 09YW15	09YW15 melanoplus
44	65	11.5	251	2 09ZEY5	09ZEY5 buchiera ap
45	65	11.5	279	2 09RC44	09RC44 streptococc

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	534 AA.
09NMU9	09NMU9			
AC	09NMU9			
DI	01-OCT-2000 (TReMBLrel. 15, Created)			
DI	01-OCT-2000 (TReMBLrel. 15, Last sequence update)			
DE	CDNA FLJ10368 FIS. CLONE NT2RM2001544.			
CS	Homo sapiens (Human).			
OC	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:			
OC	Mammalia: Eultheria: Primates: Catarrhini: Homiidae: Homo.			
OX	NCBI_TaxID:9606;			
PN	[1]			
PP	SEQUENCE FROM N.A.			
PA	Isooai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
PA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,			
PA	Watanabe M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,			
PA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,			
PA	Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuo Y.,			
PA	Ninomiya K., Iwayanagi T.,			
RI	*NED/ human cDNA sequencing project.*			
RL	Submitted (FEF-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL:AK01230; BAA91568.1; -			
SQ	SEQUENCE 534 AA: 71425 MW: C0A9C40823F9540C CRC64:			

Query Match	100.0%;	Score 567;	DB 4;	Length 634;
Best local Similarity	100.0%;	Pred. No. 1;	le-51;	
Matches	199;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
09	1 MSLEPAINIVITPLNOLKGGITIVNVGVKFFKPPYLSKGTGDSVTVTDOTNKKLTCL 60			
DB	1 MSLEPAINIVITPLNOLKGGITIVNVGVKFFKPPYLSKGTGDSVTVTDOTNKKLTCL 60			
09	61 LFGSGEALPIYKNGDVFHRLKIOVKKETOGITSSGFASLTFEET 109			
DB	61 LFGSGEALPIYKNGDVFHRLKIOVKKETOGITSSGFASLTFEET 109			
RESULT	?			
09NMU5				

```

ID 09NMUX5 PRELIMINARY: PRT: 634 AA.
AC 09NMUX5:
DI 01-OCT-2000 (ITEMBLrel. 15, Created)
DI 01-OCT-2000 (ITEMBLrel. 15, Last sequence update)
DI 01-JUN-2001 (ITEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 71.4 KDA PROTEIN (UNKNOWN) (PROTEIN FCP MGC:10290).
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
OX NCBI_TaxID=9606;
PN 111
PP SEQUENCE FROM N.A.
PC TISSUE=PLACENTA:
RA Isegaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahari K., Masuho Y., Sasaki H.,
RA "MEDO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
PP SEQUENCE FROM N.A.
RA Isegaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahari K., Masuho Y., Sasaki H.,
RA "MEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
PP SEQUENCE FROM N.A.
PC TISSUE=SKIN, AND MELANOMA:
RA Strausberg R.:
RA Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK001935: BAA91988.1: -
DR EMBL: AK022580: BAB14110.1: -
DR EMBL: BC002923: AAH02923.1: -
SQ SEQUENCE 634 AA: 71441 MW: 123A12CABE70BC91 CFC64:

Query Match 100.0% Score 567: DB 4: Length 634:
Best Local Similarity 100.0%: Pred. No. 1e-51:
Matches 109: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1 MSLLPATNYITPLNQLKGGITVNVGVKFFRPYLSKGTGYCSVTIVDQINVKLTCL 60
DI 1 MSLLPATNYITPLNQLKGGITVNVGVKFFRPYLSKGTGYCSVTIVDQINVKLTCL 60
DB 1 MSLLPATNYITPLNQLKGGITVNVGVKFFRPYLSKGTGYCSVTIVDQINVKLTCL 60
OY 61 LFSGNVEALPIIYKNGDIVFHRKLIQVYKKEGTGITSSEGFASITFEET 109
DI 61 LFSGNVEALPIIYKNGDIVFHRKLIQVYKKEGTGITSSEGFASITFEET 109
DB 61 LFSGNVEALPIIYKNGDIVFHRKLIQVYKKEGTGITSSEGFASITFEET 109

RESULT 3 PRELIMINARY: PRT: 113 AA.
OY 09CYST7:
DI 01-JUN-2001 (ITEMBLrel. 17, Created)
DI 01-JUN-2001 (ITEMBLrel. 17, Last sequence update)
DI 01-JUN-2001 (ITEMBLrel. 17, Last annotation update)
DE 2810458H15R1K PROTEIN.
GN 2810458H15R1K.
OS Mus musculus (Mouse).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
OX NCBI_TaxID=10090;
PN 111
PP SEQUENCE FROM N.A.
PC STRAIN=C57BL/6J; TISSUE=EMBRYO:
RX MEDLINE=21085650; PubMed=11217891:
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Mishi K., Kiyosaka H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Bateman A., Casavant T.,

```

```

FA Fleischer W., Gaasterland T., Gissi C., King B., Kochiwa H.,
FA Kuehl P., Lewis S., Matsuo Y., Nikaide J., Pesole G., Quackenbush J.,
FA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
FA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,
FA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
FA Brownstein M.J., Butt C., Fletcher C., Fujita M., Gariboldi M.,
FA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
FA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
FA Nordone P., Ping R., Ringwald M., Rodriguez I., Sakamoto N.,
FA "Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
FA Suzuki H., Togo-Oka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,
FA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
FA Hasehizaki T.,
FA "Functional annotation of a full-length mouse cDNA collection."
PL Mature 493,685-690(2001).
DR EMBL: AK013364: BAB28810.1: -
DR MGD: MGI:1920085: 2810458H15R1K.
SQ SEQUENCE 113 AA: 12740 MW: 08BB99FC5D0C7088 CFC64:

Query Match 57.1% Score 324: DB 11: Length 113:
Best Local Similarity 73.3%: Pred. No. 7.2e-27:
Matches 63: Conservative 6: Mismatches 17: Indels 0: Gaps 0:

OY 1 MSLLPATNYITPLNQLKGGITVNVGVKFFRPYLSKGTGYCSVTIVDQINVKLTCL 60
DI 1 MSLLPATNYITPLNQLKGGITVNVGVKFFRPYLSKGTGYCSVTIVDQINVKLTCL 60
DB 1 MSLLPATNYITPLNQLKGGITVNVGVKFFRPYLSKGTGYCSVTIVDQINVKLTCL 60
OY 61 LFSGNVEALPIIYKNGDIVFHRKLI 86
DI 61 LFSGNVEALPIIYKNGDIVFHRKLI 86
DB 61 LFSGNVEALPIIYKNGDIVFHRKLI 86

RESULT 4 PRELIMINARY: PRT: 555 AA.
OY 013988:
DI 01-JAN-1999 (ITEMBLrel. 09, Created)
DI 01-MAY-2000 (ITEMBLrel. 13, Last sequence update)
DI 01-MAY-2000 (ITEMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 64.1 KDA PROTEIN C26H5.06 IN CHROMOSOME 1.
GN SPAC226H5.06.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota: Fungi: Ascomycota: Schizosaccharomycetes:
OC Schizosaccharomycetales: Schizosaccharomycetaceae:
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
PN 111
PP SEQUENCE FROM N.A.
PC STRAIN=972:
RA Oliver K., Harris D., Bartell B.G., Rajandream M.A., Wood V.,
RA Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z52126: CAB16192.2: -
DR Hypothetical protein.
SQ SEQUENCE 555 AA: 64111 MW: A79DNA95A0C4F803 CFC64:

Query Match 20.1% Score 114: DB 3: Length 555:
Best Local Similarity 34.4%: Pred. No. 0.00074:
Matches 31: Conservative 19: Mismatches 35: Indels 6: Gaps 3:

OY 12 KESTIVNVGVKFFRPYLS-KGT-DYCSVTIVDQ-----NKLTLTLLFSGNVEALPI 71
DI 12 KESTIVNVGVKFFRPYLS-KGT-DYCSVTIVDQ-----NKLTLTLLFSGNVEALPI 71
DB 12 KESTIVNVGVKFFRPYLS-KGT-DYCSVTIVDQ-----NKLTLTLLFSGNVEALPI 71
OY 38 KKEIIVNLCGVKWDTPSRQSLHGTRKDWVTIYVMDPCTCDISIGLQIHLSKQGNLDPV 97
DI 38 KKEIIVNLCGVKWDTPSRQSLHGTRKDWVTIYVMDPCTCDISIGLQIHLSKQGNLDPV 97
DB 38 KKEIIVNLCGVKWDTPSRQSLHGTRKDWVTIYVMDPCTCDISIGLQIHLSKQGNLDPV 97

RESULT 5 PRELIMINARY: PRT: 497 AA.
OY 076380:
DI 076380:
DB 076380:

```

AC 076380: 01-NOV-1998 (TREMblrel. 08, Created)
DI 01-NOV-1998 (TREMblrel. 08, last sequence update)
DI 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE ALPHA TELOMER E BINDING PROTEIN.
OS Oxytricha tritallax.
OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Stichotrichida;
Oxytrichidae; Oxytricha.
OX NCBI_TaxID=5945;
RN [1]
RP SEQUENCE FROM N.A.
RA Prescott J.D., Dubois M.L., Prescott D.M.:
RT "Oxytricha tritallax macronuclear alpha telomere binding protein
gene".
PL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DE EMBL: AF067931; AAC7515.1; -
DR HSSP: P29549; JOTC
DR InterPro: IPR003415; Telo_bind_alpha.
DR Pfam: PF02307; Telo_bind_alpha.1.
SO SEQUENCE 497 AA: 55351 MW: 13207685633AB5665 CRC64;

Query Match 18.7% Score 105; DB 5; Length 477;
Best Local Similarity 30.4%; Pred. No. 0.0045;
Matches 31: Conservative 15; Mismatches 42; Indels 14; Gaps 2;
OY 7 TNYITFLN--OLKSGIVNYGVKFFRPYLSKGTGYCSVVIIVD-----Q 52
DB 34 SKYEVELTKADLTSTVADHFAVAVDAITFPYKINDEYICSLKIVDPYLKKEKGID 93
OY 53 TWKLTCLFSGNVEALPIYKNGDIVRFRLKIOYKKETO 94
DB 94 NSDYATLVYAKRFEDLPILHRLGDIIRHRAITRLYNGORQ 135

RESULT 6
ID 09S37 PRELIMINARY: PRT: 318 AA.
AC 09S37: 01-MAY-2000 (TREMblrel. 13, Created)
DI 01-MAY-2000 (TREMblrel. 13, last sequence update)
DI 01-MAY-2000 (TREMblrel. 13, last annotation update)
DE ATG05210. PROTEIN.
GN ATG05210.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CV. COLUMBIA:
RC MEDLINE=20083487; PubMed=10617197;
EX Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Jock C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.F., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.H., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.F., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Cressy T.H., Goodman H.M., Sorensen C.P.,
RA Coppenhaver G.P., Preuss D., Nieman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.:
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
PL Nature 402:761-768(1999).
DR EMBL: AC007018; AAD29059.1; -
SO SEQUENCE 318 AA: 35220 MW: EFFIADN077B93A2E CRC64;

Query Match 15.0% Score 85; DB 10; Length 318;
Best Local Similarity 30.4%; Pred. No. 0.45;
Matches 21: Conservative 15; Mismatches 30; Indels 2; Gaps 1;
OY 23 VNYGVKFFRPYLSKGTGYCSVVIIVD--NKLTLFSGRREALPIYKNGDIVR 80
DB 24 VELIGVLEORFPPKPCRNNDWICITRIIDIVPSFGLIVNFSKILEOLFQKHNHDMIL 83

OY 81 FRLKIOY 96
DB 84 FTRIMQIF 92

RESULT 7
ID 09Y19 PRELIMINARY: PRT: 473 AA.
AC 09Y19:
DI 01-NOV-1999 (TREMblrel. 12, Created)
DI 01-NOV-1999 (TREMblrel. 12, last sequence update)
DI 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE REPLICATION PROTEIN A LARGE SUBUNIT.
GN RPA1.
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiales; Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=KSU-1;
RA Zhu G., Marchevka M.J., Keithly J.S.:
RT "Cryptosporidium parvum has a short-type replication protein A (RPA)
large subunit".
PL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DE EMBL: AF132307; AAD42052.1; -
DR InterPro: IPR00409; P1_PLC_X.
DR PROSITE: PS50007; P1PLC_X_DOMAIN: 1.
SO SEQUENCE 473 AA: 53997 MW: DEC281A982343878 CRC64;

Query Match 13.7% Score 77.5; DB 5; Length 473;
Best Local Similarity 23.0%; Pred. No. 4.4;
Matches 26: Conservative 27; Mismatches 51; Indels 9; Gaps 4;
OY 2 SLVPAINTYIPLNKG---GIVNYGVKFFRP---YLSKGTGYCS--VVIIVDQI 53
DB 111 SGKAPKIVFTIVKDIPEIPVPSIVDLIGIVHFSPTKVISRKNNDSVSRITSVDKI 170
OY 54 NKLTCLFSGNVEALPIYKNGDIVRFRLKIOYKKETOIGTSGFASLT 106
DB 171 GFLNITLWGLAEVSDKFLGNPVVALKSTOIRRYQOR--GGSTLNGSNNEF 222

RESULT 8
ID 09C28 PRELIMINARY: PRT: 759 AA.
AC 09C28:
DI 01-JUN-2001 (TREMblrel. 17, Created)
DI 01-JUN-2001 (TREMblrel. 17, last sequence update)
DI 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE CONSERVED HYPOTHEICAL PROTEIN.
GN 3H10.40.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariiales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hohelsel J., Brandt P., Farlmann B., Holland R.,
RA Nykatura G., Maves H.W., Mannhaupt G.:
RT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
PL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL513442; CAC28643.1; -
SO SEQUENCE 759 AA: 85218 MW: DC737A5574FF4F73 CRC64;

Query Match 13.2% Score 75; DB 3; Length 759;
Best Local Similarity 25.0%; Pred. No. 14;

```

Matches 27: Conservative 24: Mismatches 41: Indels 16: Gaps 4:
OY 3 LVPATNYIPIPLNOL-----KGIIVVGVVKKFFPKPYISKGDYCSVTI-----VD 51
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 18 LFPST---LIPRALLDNDGDADPGSMVAVICVLKDCAPVAVIHSMDKCLITLISLSE 74
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 52 OTNRLTCLLFGSNGYFALFIIRKNDIVRFHRLKIOVKKEIFGSISS 99
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 75 DESAGVELVIF--PPDARPEVAGADVLVLSAKVQPFKSNFKSLITS 120
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 9
O9CE7 PRELIMINARY: PRI: 383 AA.
ID O9CE7
AC O9CE7
DI 01-JUN-2001 (FIREBLREL: 17, Created)
DI 01-JUN-2001 (FIREBLREL: 17, Last sequence update)
DI 01-JUN-2001 (FIREBLREL: 17, Last annotation update)
DE 231001SEI4RIK PROTEIN.
CN 231001SEI4RIK
OS Mus musculus (Mouse).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
CX NCBI_TaxID=10090:
EN 11
PP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD, AND TONGUE;
RX MEDLINE=21085650; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii T.,
RA Arakawa T., Hara A., Fukumishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito T.,
RA Kadota K., Matsuda H.A., Asburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA Kuenli P., Lewis S., Matsuura Y., Nikaido I., Pesole G., Ouellette J.,
RA Schiml L.M., Stabili F., Suzuki R., Tomita M., Wagner L., Wasth J.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brustein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
PA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee S.H.,
PA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts J.,
PA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
PA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.F.,
PA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilting L.,
PA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kchitsuki S.,
PA Hayashizaki Y.:
PI *Functional annotation of a full-length mouse cDNA collection.
PI Nature 409:585-590(2001).
PI EMBL: AK013942; BAB2027.1: -.
PI EMBL: AK003350; BAB2023.1: -.
PI MGD: MGI:1913615; 231001SEI4RIK.
SO SEQUENCE 383 AA; 43208 MW; 97EF08712201E0C2 CRC64:

Query Match 13.1% Score 74.5; DB 11: Length 383;
Best Local Similarity 24.4% Pred. No. 7.2;
Matches 19: Conservative 21: Mismatches 29: Indels 9: Gaps 3:
OY 23 VVVYGVVKKFFPKPYISKGDYCSVTIIVDOIVNKL--TCLLFGSNGYFALFIIRK--R 78
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 240 INVTHYIK-----HLSFGEDIPGIVNPLDHNIVTAPOASMPFQYFKVYFIVMKV 254
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 79 VRFHRLKIOVKKEITGI 95
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 295 LRTNPFVTRPEKXVANGL 312
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 10
O9CE7 PRELIMINARY: PRI: 198 AA.
ID O9CE7
AC O9CE7
DI 01-MAY-1999 (FIREBLREL: 10, Created)
DI 01-MAY-1999 (FIREBLREL: 10, Last sequence update)
DI 01-MAY-1999 (FIREBLREL: 10, Last annotation update)

```

```

DI 01-MAY-1999 (FIREBLREL: 10, Last annotation update)
DE HYPOHETICAL 22.4 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.
CX NCBI_TaxID=9606:
EN 11
PP SEQUENCE FROM N.A.
RA Barrow J.K.-P., Boguski M.S., Touchman J.J., Spencer F.,
PI *Full-insert sequence of mapped XREF EST.
PI Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DE EMBL: AF091085; AAC72954.1: -.
KW Hypothetical protein
SO SEQUENCE 198 AA; 22396 MW; 75CE7A231C91EF02 CRC64:

Query Match 13.1% Score 74; DB 4: Length 198;
Best Local Similarity 26.2% Pred. No. 3.7;
Matches 17: Conservative 18: Mismatches 26: Indels 4: Gaps 2:
OY 36 YISKGDYCSVTIIVDOIVNKL--TCLLFGSNGYFALFIIRK--NGDIVRFHRLKIOVKK 91
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 111 HLTSGEDIPGIVNPLDHNIVTAPOASMPFQYFKVAVPTVYKVDGVLRTNPFVTRHEK 170
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 32 ETGGI 95
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 171 VANGL 175
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 11
O9CE7 PRELIMINARY: PRI: 228 AA.
ID O9CE7
AC O9CE7
DI 01-JUN-2001 (FIREBLREL: 17, Created)
DI 01-JUN-2001 (FIREBLREL: 17, Last sequence update)
DI 01-JUN-2001 (FIREBLREL: 17, Last annotation update)
DE D347704.2.2 (CGI-54 (LOC51614), ISOFORM 2) (FRAGMENT).
GN D347704.2.
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominiidae: Homo.
CX NCBI_TaxID=9606:
EN 11
PP SEQUENCE FROM N.A.
RA Brown A.:
PI Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DE EMBL: AL121586; rAC33497.1: -.
PI NON-ITER
SO SEQUENCE 228 AA; 25918 MW; 97888251C9A7B93 CRC64:

Query Match 13.1% Score 74; DB 4: Length 228;
Best Local Similarity 26.2% Pred. No. 4.4;
Matches 17: Conservative 18: Mismatches 26: Indels 4: Gaps 2:
OY 36 YISKGDYCSVTIIVDOIVNKL--TCLLFGSNGYFALFIIRK--NGDIVRFHRLKIOVKK 91
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 139 HLTSGEDIPGIVNPLDHNIVTAPOASMPFQYFKVAVPTVYKVDGVLRTNPFVTRHEK 198
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
OY 32 ETGGI 95
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 199 VANGL 203
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 12
O9CE7 PRELIMINARY: PRI: 346 AA.
ID O9CE7
AC O9CE7
DI 01-JUN-2000 (FIREBLREL: 15, Created)
DI 01-JUN-2000 (FIREBLREL: 15, Last sequence update)
DI 01-JUN-2000 (FIREBLREL: 15, Last annotation update)

```

```

OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
OX NCBI_TaxID=9606:
RN
RP SEQUENCE FROM N.A.
PC TISSUE=LIVER:
RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,
RA Liu M., He F.:
RT *Functional prediction of the coding sequences of 121 new genes
RT deduced by analysis of cDNA clones from human fetal liver.*
PL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF115614; AAF71038.1;
SO SEQUENCE 345 AA: 39228 MW: A04222C998834BF CRC64:

Query Match 13.1% Score 74: DB 4: Length 345:
Best Local Similarity 26.2% Pred. No. 7.2:
Matches 17: Conservative 18: Mismatches 26: Indels 4: Gaps 2:

OY 36 YLSKGDICSVTVTDQINVKL--TCLFSGNVEALPIYK--NGDIVRFHRLKIVYKK 91
DB 211 HLSFGEDYPGIVPLDHTNVIAPOASMMGFYFVYKVPYVYKVDGEVLRITQFSVTPHEK 270
OY 92 ETGGI 36
DB 271 VANGI 275

RESULT 13
OY282 PRELIMINARY: PRI: 383 AA.
ID O9Y282:
AC O9Y282:
DI 01-NOV-1999 (IREMBLrel. 12, Created)
DI 01-NOV-1999 (IREMBLrel. 12, Last sequence update)
DI 01-MAR-2001 (IREMBLrel. 16, Last annotation update)
DE CGI-54 PROTEIN (DJ47704.2) (CGI-54).
GN DJ47704.2.
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
OX NCBI_TaxID=9606:
RN
RP SEQUENCE FROM N.A.
PC Lih W.-C.:
RT *Comparative gene cloning: Identification of novel human genes with
RT Caenorhabditis elegans proteome as template.*
PL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Zhang J., Liu T., Ye M., Zhang Q., Fu G., Zhou J., Wu J., Shen Y.,
RA Yu M., Chen S., Mao M., Chen Z.:
RT *Human hypothetical 43.2 kd protein.*
PL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Brown A.:
PL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF151812; AAD34049.1;
DR EMBL: AF077030; AAD27763.1;
DR EMBL: AL121586; CAB89412.1;
SO SEQUENCE 383 AA: 43222 MW: 8370FB3067AD95CE CRC64:

Query Match 13.1% Score 74: DB 4: Length 383:
Best Local Similarity 26.2% Pred. No. 8.1:
Matches 17: Conservative 18: Mismatches 26: Indels 4: Gaps 2:

OY 36 YLSKGDICSVTVTDQINVKL--TCLFSGNVEALPIYK--NGDIVRFHRLKIVYKK 91
DB 248 HLSFGEDYPGIVPLDHTNVIAPOASMMGFYFVYKVPYVYKVDGEVLRITQFSVTPHEK 307
OY 92 ETGGI 96

```

```

DB 308 VANGI 312

RESULT 14
OY281 PRELIMINARY: PRI: 299 AA.
ID OY281:
AC OY281:
DI 01-OCT-2000 (IREMBLrel. 15, Created)
DI 01-OCT-2000 (IREMBLrel. 15, Last sequence update)
DI 01-OCT-2000 (IREMBLrel. 15, Last annotation update)
DE BACN4124.C.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota: Metazoa: Arthropoda: Insecta:
OC Echinoidea: Drosophilidae: Drosophila.
OX NCBI_TaxID=7227:
RN
RP SEQUENCE FROM N.A.
RA Moller, Gadieu, Dreano, Lelaure, Galibert F.:
RT *Sequencing the distal X chromosome of Drosophila melanogaster.*
PL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Bence F.:
PL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL121805; CAB58072.1;
SO SEQUENCE 299 AA: 32007 MW: C7224AFCDC0809B6 CRC64:

Query Match 12.8% Score 72.5: DB 5: Length 299:
Best Local Similarity 33.3% Pred. No. 8.8:
Matches 19: Conservative 14: Mismatches 19: Indels 5: Gaps 1:

OY 52 QIVNKLICILFSGNVEALPIYK----NGDIVRFHRLKIVYKKITGTCSTGAS 103
DB 21 KNSLLITLLAILHLENSKIVKPKPLGNSNITIKRKKIKPKVETSTMSSTGAS 77

RESULT 15
OY28X PRELIMINARY: PRI: 937 AA.
ID OY28X:
AC OY28X:
DI 01-OCT-2000 (IREMBLrel. 15, Created)
DI 01-OCT-2000 (IREMBLrel. 15, Last sequence update)
DI 01-JUN-2001 (IREMBLrel. 17, Last annotation update)
DE F2225.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
OX NCBI_TaxID=3702:
RN
RP SEQUENCE FROM N.A.
RA Shing P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altaji H., Rei O., Ching C., Chiu J., Choi E., Conn L.,
RA Conkey A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Leng C., Li J., Liu A., Liu K., Liu S., Mukhariskiy N., Nguyen M.,
RA Poirier M., Phan P., Sakano H., Schwartz J., Southwick A., Thayer A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.:
RT *Genetic sequence for Arabidopsis thaliana BAC F2205 from chromosome
RT 1.*
PL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.:
PL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.:
PL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 12, 2002, 08:33:16 ; Search time 12.47 Seconds

(without alignments)
195.701 Million cell updates/sec

Title: US-09-816-248-5

Sequence: 1 MSLEPAINVIYPLNOLKGC.....KKEIGITSSGASLIFEET 109

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5
212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summariesDatabase : Issued Patents AA: *
1: /cgn2_5/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_5/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_5/ptodata/2/1aa/6A.COMB.pep: *
4: /cgn2_5/ptodata/2/1aa/5B.COMB.pep: *
5: /cgn2_5/ptodata/2/1aa/PTUS.COMB.pep: *
6: /cgn2_5/ptodata/2/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70.5	12.4	287	3	US-08-733-230-2
2	70.5	12.4	287	4	US-08-953-326-2
3	64.5	11.4	412	2	US-08-741-134-2
4	62	10.9	990	2	US-08-645-153B-15
5	61.5	10.8	155	2	US-08-162-402B-18
6	60	10.6	990	2	US-08-392-625-20
7	60	10.6	990	2	US-08-466-951A-20
8	59	10.4	405	1	US-07-996-772A-4
9	59	10.4	405	1	US-08-445-822-4
10	59	10.4	405	5	PCT-US93-12585-4
11	58	10.2	85	2	US-08-480-229C-5
12	58	10.2	1169	1	US-08-659-235C-5
13	58	10.2	1169	2	US-08-542-921-2
14	58	10.2	1169	2	US-08-880-685-2
15	58	10.2	1169	2	US-08-880-684-2
16	57.5	10.1	935	2	US-08-152-721B-2
17	57.5	10.1	383	2	US-08-569-168-7
18	57	10.1	387	1	US-07-996-772A-2
19	57	10.1	387	1	US-08-446-822-2
20	57	10.1	387	5	PCT-US93-12586-2
21	56.5	10.0	504	1	US-07-932-915-2
22	56.5	10.0	504	1	PCT-US91-05825-2
23	56.5	10.0	542	4	US-09-189-462-6
24	56.5	10.0	906	1	US-08-220-151-9
25	56.5	10.0	906	1	US-08-473-445-2
26	56.5	10.0	907	1	US-08-349-006-2
27	56.5	10.0	907	1	US-08-349-006-2

28	56.5	10.0	907	3	US-08-804-439A-19	Sequence 19, App1
29	56.5	10.0	907	3	US-08-720-229-19	Sequence 19, App1
30	56.5	10.0	907	5	PCT-US94-04180-2	Sequence 21, App1
31	56	9.9	331	2	US-08-878-989-21	Sequence 21, App1
32	56	9.9	331	3	US-09-101-146-64	Sequence 64, App1
33	56	9.9	331	4	US-09-272-795-21	Sequence 21, App1
34	56	9.9	192	1	US-07-783-705A-4	Sequence 4, App1
35	55.5	9.8	863	4	US-09-238-303-11	Sequence 11, App1
36	55.5	9.8	592	3	US-08-991-813-2	Sequence 2, App1
37	54.5	9.6	213	2	US-08-716-284-2	Sequence 2, App1
38	54.5	9.6	251	1	US-08-425-336-100	Sequence 100, App
39	54.5	9.6	251	1	US-08-425-336-100	Sequence 101, App
40	54.5	9.6	251	1	US-08-488-113B-100	Sequence 101, App
41	54.5	9.6	251	1	US-08-488-113B-101	Sequence 101, App
42	54.5	9.6	251	1	US-08-477-484B-100	Sequence 100, App
43	54.5	9.6	251	1	US-08-477-484B-101	Sequence 101, App
44	54.5	9.6	251	2	US-08-646-360-100	Sequence 100, App
45	54.5	9.6	251	2	US-08-646-360-101	Sequence 101, App

ALIGNMENTS

RESULT 1
US-08-733-230-2
: Sequence 2, Application US/08733230
: Patent No. 6025338
GENERAL INFORMATION:
: APPLICANT: Barbet, Anthony F.
: APPLICANT: Ganla, Roman Reddy
: APPLICANT: McGuire, Travis C.
: APPLICANT: Burridge, Michael J.
: APPLICANT: Nylka, Aceme
: APPLICANT: Rurangirwa, Fred R.
: APPLICANT: Mahan, Sunan M.
: TITLE OF INVENTION: Nucleic Acid Vaccines Against
: TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Salivanchik & Salivanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: USA
: ZIP: 32606
COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/733-230
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Whitlock, Ted W.
: REGISTRATION NUMBER: 36,965
: REFERENCE/Docket NUMBER: UF-167
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 352-375-8100
: TELEFAX: 352-372-5800
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 287 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-733-230-2

Query: Match 12.4% Score 70.5; DB 3; Length 287;
Best Local Similarity 30.3%; Pred. No. 0.47;
Matches 27; Conservative 19; Mismatches 18; Indels 25; Gaps 7;

[illegible]

```

: RESULT 2
: US-08-953-326-2
: Sequence 2, Application US/08553326
: Patient No. 6251872
: GENERAL INFORMATION:
: APPLICANT: Barbel, Anthony F.
: APPLICANT: Ganta, Poman R.
: APPLICANT: McGuire, Travis C.
: APPLICANT: Burrigdge, Michael J.
: APPLICANT: Nyika, Aceme
: APPLICANT: Purangirwa, Fred R.
: APPLICANT: Mahan, Susan M.
: TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
: TITLE OF INVENTION: Animals and Humans
: FILE REFERENCE: UF-16701
: CURRENT APPLICATION NUMBER: US/08/953.326
: CURRENT FILING DATE: 1997-10-17
: EARLIER APPLICATION NUMBER: 08/953.326
: EARLIER FILING DATE: 1997-10-17
: EARLIER APPLICATION NUMBER: 08/733.230
: EARLIER FILING DATE: 1995-10-17
: NUMBER OF SEQ ID NOS: 24
: SOFTWARE: Patentl Ver. 2.0
: SEQ ID NO 2
: LENGTH: 287
: TYPE: PRT
: ORGANISM: Coecidia ruminantium
: US-08-953-326-2

```

Query Match	12.4%	Score 70.5	DB 4	Length 287
Best Local Similarity	30.3%	Pred. No. 0.47		
Matches	27	Conservative	19	Mismatches 18; Indels 25; Gaps 7
Oy	35	PLYSK--GDYCSYIVDQIVNWKTC-----LFSGYNELPIYIKRQDVIHPRL---84		
Db	194	PLYCAAGIGD---LVSIVAIHAFKSTQSGKIGISTISNSEA--SIFIG3--HPRVIGN ²⁴⁵		
Oy	85	-----KIDVYKKEQTGISTSGFASLTTE 107		
Db	246	EFKDIALIKLIFTSK1-GISNPGFASALTD 273		

RESULT 3
 US-08-741-134-2
 : Sequence 2, Application US/08741134
 : Patent No. 5851498
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Litack, Gerald
 : APPLICANT: Altmerrl, Enad S.
 : APPLICANT: Fernandes, Altmerrl, Teresa
 : TITLE OF INVENTION: IMMUNOPHILIN FRBP45 AND COMPOSITIONS FOR MAKING
 : TITLE OF INVENTION: AND
 : TITLE OF INVENTION: METHODS OF USING THE SAME
 : NUMBER OF SEQUENCES: 6
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5851498RIS
 : STREET: One Liberty Place - 45th floor
 : CITY: Philadelphia
 : STATE: Pennsylvania
 : COUNTRY: USA
 : ZIP: 19103
 :
 : COMPUTER READABLE FORM:

```

1 MEDIUM TYPE: floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: Windows 3.11
4 SOFTWARE: WordPerfect for Windows
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/08/741,134
7 FILING DATE:
8 CLASSIFICATION: 424
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: US 60/007,163
11 FILING DATE: 01-NOV-1995
12 CLASSIFICATION: 424
13 ATTORNEY/AGENT INFORMATION:
14 NAME: DeLuca, Mark
15 REGISTRATION NUMBER: 33,229
16 REFERENCE/DOCKET NUMBER: TJU-20900
17 TELECOMMUNICATION INFORMATION:
18 TELEPHONE: 215-566-3100
19 TELEFAX: 215-568-3439
20 INFORMATION FOR SEQ ID NO: 2:
21 SEQUENCE CHARACTERISTICS:
22 LENGTH: 412 amino acids
23 TYPE: amino acid
24 TOPOLOGY: linear
25 MOLECULE TYPE: protein
26 US-08-741-134-2

```

```
Query Match:          11.4%   Score 64.5; DB 2: Length 412;
Best Local Similarity 32.2%; Pred No. 4.6;
Matches    12; Conservative    9; Mismatches    20; Indels    11; Gaps    3;
```

QY 33 KPFPLKNG-----TDYCSVTIIVDITNWKTCLEFSGNYELP--IIYKKGIIVRF 81
 *:::1:::1:::1:::1:::1:::1:::1:::1:::1:::1:::1:::1:::
DB19 KPFHSIAAMADISTGNDPCGVAVNVVDAKNR-LVCTTLOKGNI IOVPDLIFKSGDSVSF 75

RESULT 4
US-08-545-1938-15
Sequence 15, Application US/08645193B
Patent No. 5965253
GENERAL INFORMATION:
APPLICANT: Kupke, Thomas
APPLICANT: Golz, Friedrich
APPLICANT: Kempler, Christoph
APPLICANT: Jung, Gunther
TITLE OF INVENTION: Oxidative Decarboxylation of Peptides
TITLE OF INVENTION: Catalyzed by Flavoprotein Epub
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
Zip: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,193B
FILING DATE: 13-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0652.1540000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:
 LENGTH: 990 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 US-08-645-1938-15

Query Match: 10.9% Score 62; DB 2: Length 990;
 Best Local Similarity 23.9%; Pred. No. 32;
 Matches 28: Conservative 18; Mismatches 41; Indels 30; Gaps 6;

QY 4 VPAT-----NYITPLNOLKGTIVNVGVK-----FFKPPYLSKGTIDYCSVTIV 39
 Db 735 IPKTYDNFTIDYLLPITTELKVNINFKFFYIKFKEDEDFIKRLRLREDEDEYSQIVSFI 743

OY 51 DOTWKLICLFSGNY-----EALPIIYKNG-----DIVFRH----LKIVVKKKE 92
 Db 795 --KMKKQYCLINSELVDSYDVYEVYRIGGPHVIEDIEFMWDSLSIHIIOSE 849

RESULT 5
 US-08-162-402B-18
 : Sequence 18, Application US/08152402B
 : Patent No. 5972337
 : GENERAL INFORMATION:
 : APPLICANT: CERIANI, ROBERTO L.
 : APPLICANT: PETERSON, JERRY A.
 : APPLICANT: LARocca, DAVID J.
 : TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
 : NUMBER OF SEQUENCES: 29
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Pretly, Schroeder & Poplowski
 : STREET: 444 South Flower St., 19th floor
 : CITY: Los Angeles
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 90071
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: DOS
 : SOFTWARE: FastSeq for Windows Version 2.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/152.402B
 : FILING DATE: 03-DEC-1993
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER:
 : FILING DATE:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Amzel, Viviana
 : REGISTRATION NUMBER: 30,930
 : REFERENCE/DOCKET NUMBER: P65 38215
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 213-622-7700
 : TELEFAX: 213-489-4210
 : TELEX:
 : INFORMATION FOR SEQ ID NO: 18:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 156 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: unknown
 : TOPOLOGY: unknown
 : MOLECULE TYPE: peptide
 : US-08-162-402B-18

Query Match: 10.8% Score 61.5; DB 2: Length 155;
 Best Local Similarity 23.3%; Pred. No. 2.3;
 Matches 24: Conservative 12; Mismatches 27; Indels 19; Gaps 4;

OY 29 VKFFPPYLSKGTIDYCSVTIVADOTWKLICLFSGNVEALPIIYK---NGDIVRFHRLK 85
 Db 82 VKSYKVOISNGDEM---IILKDNK-----HLVGTGNIDATDVYRFSKPVITIRFRLR 134
 OY 86 IOVKKKEIGTITSSGFASLTFE 107
 Db 135 -----PVIMENGISLTFE 147

RESULT 5
 US-08-392-625-20
 : Sequence 20, Application US/08392625
 : Patent No. 5837485
 : GENERAL INFORMATION:
 : APPLICANT: Enliam, Karl-Dieter
 : APPLICANT: Gtz, Friedrich
 : APPLICANT: Schnell, No. 5837485bert
 : APPLICANT: Augustin, Johannes
 : APPLICANT: Engelke, Gernar
 : APPLICANT: Rosenstein, Ralf
 : APPLICANT: Kalotta, Corlina
 : APPLICANT: Klein, Cora
 : APPLICANT: Wieland, Bernd
 : APPLICANT: Kupke, Thomas
 : APPLICANT: Jung, G. ncher
 : APPLICANT: Kellner, Roland
 : TITLE OF INVENTION: Biosynthetic Process For The Preparation
 : NUMBER OF SEQUENCES: 42
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 : STREET: 1100 New York Avenue
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: U.S.A.
 : ZIP: 20005
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/392.625
 : FILING DATE:
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/876,791
 : FILING DATE: 30-APR-1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Esmond, Robert W.
 : REGISTRATION NUMBER: 32,893
 : REFERENCE/DOCKET NUMBER: 0552.0980002
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (202) 371-2500
 : TELEFAX: (202) 371-2540
 : INFORMATION FOR SEQ ID NO: 20:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 990 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : US-08-392-625-20

Query Match: 10.6% Score 60; DB 2: Length 990;
 Best Local Similarity 24.5%; Pred. No. 59;
 Matches 25: Conservative 16; Mismatches 35; Indels 26; Gaps 5;

OY 4 VPAT-----NYITPLNOLKGTIVNVGVK-----FFKPPYLSKGTIDYCSVTIV 50
 Db 734 IPKTYDNFTIDYLLPITTELKVNINFKFFYIKFKEDEDFIKRLRLREDEDEYSQIVSFI 793
 OY 51 DOTWKLICLFSGNY-----EALPIIYKNG-----DIVFR 81

Db 794 --KMKDYCLNSELVDYSIVDYVEVRYGCPHYIEDIENF 833

RESULT 7

US-08-466-961A-20
: Sequence 20, Application US/08466961A
: Patent No. 5843709
: GENERAL INFORMATION:
: APPLICANT: Entian, Karl-Dieter
: APPLICANT: Gitz, Friedrich
: APPLICANT: Schnell, No. 5843709bert
: APPLICANT: Augustin, Johannes
: APPLICANT: Engelke, Gernar
: APPLICANT: Rosenstein, Ralf
: APPLICANT: Kaletta, Corina
: APPLICANT: Klein, Cora
: APPLICANT: Wieland, Bernd
: APPLICANT: Kupke, Thomas
: APPLICANT: Jung, G nther
: APPLICANT: Kellner, Poland
: TITLE OF INVENTION: Biosynthetic Process for the Preparation of
: TITLE OF INVENTION: Chemical Compounds
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox
: STREET: 1100 New York Avenue, NW
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/466,961A
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/392,625
: FILING DATE: 22-FEB-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/876,791
: FILING DATE: 30-APR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/784,234
: FILING DATE: 31-OCT-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Esmond, Robert W.
: REGISTRATION NUMBER: 32,893
: REFERENCE/DOCKET NUMBER: 0652-0980004
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 990 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: US-08-466-961A-20

Query Match 10.6% Score 60: DB 2: Length 990:
Best Local Similarity 24.5% Pred. No. 59:
Matches 25: Conservative 16: Mismatches 35: Indels 25: Gaps 5:

OY 4 VPAI-----NYITPLNOLKSGTIVNYGVK-----FFKPYLSGTDYCSVTIV 60
Db 734 IPTIYDNEIYDILLPFIILKYNNEFFYIKREDEDFILRLFELEDYDSQIVSFI 733
OY 51 DOJNVLTLFLFSGRY-----EALPIYKNG-----DIVPF 81

Db 794 --KMKDYCLNSELVDYSIVDYVEVRYGCPHYIEDIENF 833

RESULT 8

US-07-996-772A-4
: Sequence 4, Application US/07996772A
: Patent No. 5472866
: GENERAL INFORMATION:
: APPLICANT: Gerald, Christophe
: APPLICANT: Hartig, Paul R.
: APPLICANT: Branchek, Theresa A.
: APPLICANT: Weinschenk, Richard L.
: TITLE OF INVENTION: DNA ENCODING 5-HT4a SEROTONIN
: TITLE OF INVENTION: RECEPTORS AND USES THEREOF
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: COOPER & DUNHAM
: STREET: 30 ROCKEFELLER PLAZA
: CITY: NEW YORK
: STATE: NEW YORK
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/996,772A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: White, P. John
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 42667/JPW/TEP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 977-9550
: TELEFAX: (212) 664-0525
: TELEX: 422523 COOP U1
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 406 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-07-996-772A-4

Query Match 10.4% Score 59: DB 1: Length 406:
Best Local Similarity 28.4% Pred. No. 23:
Matches 23: Conservative 12: Mismatches 30: Indels 16: Gaps 5:

OY 7 TWITITPL--NOLKSGTIVNYGVKFFKPYLSKGTIDYCSVTIVDQIVKLT----- 58
Db 55 IWFIIVSLAFADLLVSVLNAFGAIEIVDQIWF-YGEMFCIVRISLD---VLLITISIFH 110
OY 59 -CLIFSGNVEAL---PIYKN 75
Db 111 LCCLSLDRIVAIICOPLYVRN 131

US-08-446-822-4

Query Match 10.4% Score 59: DB 1: Length 406:
Best Local Similarity 28.4% Pred. No. 23:
Matches 23: Conservative 12: Mismatches 30: Indels 16: Gaps 5:

OY 4 VPAI-----NYITPLNOLKSGTIVNYGVK-----FFKPYLSGTDYCSVTIV 60
Db 734 IPTIYDNEIYDILLPFIILKYNNEFFYIKREDEDFILRLFELEDYDSQIVSFI 733
OY 51 DOJNVLTLFLFSGRY-----EALPIYKNG-----DIVPF 81

```

: CITY: NEW YORK
: STATE: NEW YORK
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/446,822
: FILING DATE: June 1, 1995
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: White, P., John
: REGISTRATION NUMBER: 28,578
: REFERENCE/DOCKET NUMBER: 42667-A-PCT-US/JPW/MAT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 291-0525
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 406 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-446-822-4

Query Match      10.4% Score 59; DB 1; Length 406;
Best Local Similarity 28.4%; Pred. No. 23;
Matches 23; Conservative 12; Mismatches 30; Indels 6; Gaps 5;

OY 7 TNYITFL--NOLKSGIVNVGVKFFKPPYLSKGTDCSVTIVDQINVKLT-----56
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 55 TNYITFLADLVSLVNAFGEIIVODIMF--YGEMLVLTSTID---VLTITASTIFH 110
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 59 -CLFSGNEAL--PIIYKN 75
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 111 LCCSLDRYVAICCPLYRN 131

RESULT 10
PCT-US93-12586-4
: Sequence 4, Application PC/TUS9312586
: GENERAL INFORMATION:
: APPLICANT: SYNAPTIC PHARMACEUTICAL CORPORATION
: TITLE OF INVENTION: DNA ENCODING 5-HT4 SEROTONIN RECEPTORS
: TITLE OF INVENTION: AND USES THEREOF
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: COOPER & DUNHAM
: STREET: 30 ROCKEFELLER PLAZA
: CITY: NEW YORK
: STATE: NEW YORK
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/12586
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: White, P., John
: REGISTRATION NUMBER: 28,578
: REFERENCE/DOCKET NUMBER: 42667-A-PCT-US/JPW/IEP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 977-9550
: TELEFAX: (212) 564-0525
: TELEX: 422523 COOP UT
: INFORMATION FOR SEQ ID NO: 4:

```

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 406 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US93-12586-4

Query Match      10.4% Score 59; DB 5; Length 406;
Best Local Similarity 28.4%; Pred. No. 23;
Matches 23; Conservative 12; Mismatches 30; Indels 16; Gaps 5;

OY 7 TNYITFL--NOLKSGIVNVGVKFFKPPYLSKGTDCSVTIVDQINVKLT-----58
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 55 TNYITFLADLVSLVNAFGEIIVODIMF--YGEMLVLTSTID---VLTITASTIFH 110
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 59 -CLFSGNEAL--PIIYKN 75
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 111 LCCSLDRYVAICCPLYRN 131

RESULT 11
US-08-480-229C-5
: Sequence 5, Application US/08480229C
: Patent No. 5874562
: GENERAL INFORMATION:
: APPLICANT: Quentromus, Thomas
: APPLICANT: Hogan, Bridgid
: APPLICANT: Snodgrass, H. Ralph
: APPLICANT: Zupancic, Thomas J.
: TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
: TITLE OF INVENTION: CELL LOCUS-1
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Pennie & Edmonds LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: United States
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/480,229C
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Poissant, Brian M.
: REGISTRATION NUMBER: 28,462
: REFERENCE/DOCKET NUMBER: 8907-0026-999
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 Pennie
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 85 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: unknown
: MOLECULE TYPE: peptide
: US-08-480-229C-5

Query Match      10.2% Score 58; DB 2; Length 85;
Best Local Similarity 30.0%; Pred. No. 3;4;
Matches 18; Conservative 10; Mismatches 24; Indels 8; Gaps 2;

OY 29 YFFKPPYISGTYRCVSVITVDQINVKLTGLLFSGNEALPIIYK--NDQIVRFHRLK 85
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 27 VKSKYKDSSNGEDMIXILKGNKH-----LVFTGHTDAIDVYRPFSPVITFEVLR 81

RESULT 12

US-08-659-235C-5
Sequence 5: Application US/08659235C
Patent No. 5877281

GENERAL INFORMATION:

APPLICANT: Quartermours, Thomas
APPLICANT: Hogan, Bridgid
APPLICANT: Snodgrass, H. Ralph
APPLICANT: Zupancic, Thomas J.
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
TITLE OF INVENTION: CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/659,235C
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0034-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 859-8854/9741
TELEX: 66141 Pennie

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown

MOLECULE TYPE: peptide
US-08-659-235C-5

Query Match 10.2% Score 58; DB 2: Length 85;
Best Local Similarity 30.0%; Pred. No. 3.4;

Matches 18: Conservative 10: Mismatches 24: Indels 8: Gaps 2:

Db 27 VKSKYKDSSNGEDMIXILKGNKH-----LVFTGHTDAIDVYRPFSPVITFEVLR 81

RESULT 13

US-08-542-921-2
Sequence 2: Application US/08542921
Patent No. 5736514

GENERAL INFORMATION:

APPLICANT: IIZUKA, TOSHIHIKO
APPLICANT: TAGAWA, MICHITO
APPLICANT: ARAI, SATOSHI
APPLICANT: MIYAKE, TOSHIRO
APPLICANT: MIYAKE, TOSHIRO

TITLE OF INVENTION: NOVEL BACILLUS STRAIN AND HARMFUL
TITLE OF INVENTION: ORGANISM CONTROLLING AGENTS

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/542,921
FILING DATE: 13-OCT-1995
CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 276082/94
FILING DATE: 14-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 49-209-0

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 24855 OPAT UR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 119 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-542-921-2

Query Match 10.2% Score 58; DB 1: Length 119;
Best Local Similarity 32.4%; Pred. No. 1.3e+02;

Matches 23: Conservative 9: Mismatches 31: Indels 8: Gaps 4:

Db 35 YLSKGTGYCVTVIVDTWKLTLFSGNYEALPIYKNGDIVR-FHRLK-IOVYRKE 92

Db 439 YLSSG--QVSGISGYTOGIPAVCLQORNSDLPINFGEDJINYSRSLHITQVRFQ 455

Db 93 TCGITSSGFAS 103

Db 497 A---TQSGSPS 504

RESULT 14

US-08-890-685-2
Sequence 2: Application US/08880685
Patent No. 5834226

GENERAL INFORMATION:

APPLICANT: IIZUKA, TOSHIHIKO
APPLICANT: TAGAWA, MICHITO
APPLICANT: ARAI, SATOSHI
APPLICANT: MIYAKE, TOSHIRO
APPLICANT: MIYAKE, TOSHIRO

TITLE OF INVENTION: NOVEL BACILLUS STRAIN AND HARMFUL
TITLE OF INVENTION: ORGANISM CONTROLLING AGENTS

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

1 NAME: OBLON, NORMAN F
2 REGISTRATION NUMBER: 24,618
3 REFERENCE/DOCKET NUMBER: 49-209-0
4 TELECOMMUNICATION INFORMATION:
5 TELEPHONE: (703) 413-3000
6 TELEFAX: (703) 413-2220
7 TELEX: 248455 OVAT UR
8
9 INFORMATION FOR SEQ ID NO: 2:
10 SEQUENCE CHARACTERISTICS:
11 LENGTH: 1169 amino acids
12 TYPE: amino acid
13 TOPLOGY: linear
14 MOLECULE TYPE: protein
15
16 US-08-880-691-2

```

Query Match	10.2%	Score 58:	DB 2:	Length 1169:
Best Local Similarity	32.4%	Pred. No.	1.3e+02:	
Matches	23:	Conservative	9:	Mismatches 31:
			Indels	8:
			Gaps	4:

Q7 93 LCHITSGFAS 103
I I I I
Db 497 A---TSGSPS 504

Q7 93 LCHITSGFAS 103
I I I I
Db 497 A---TSGSPS 504

Q7 93 LCHITSGFAS 103
I I I I
Db 497 A---TSGSPS 504

Search completed: April 12, 2002, 08:34:46
Job time: 30 sec

Email: coapbs-r@mail.nih.gov
 Tissue Procurement: ATCC/DCID/DTP
 cDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 Sequencing by: National Institutes of Health Intramural
 Gaithersburg, Maryland:
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@hgrl.nih.gov
 Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
 Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J.J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
 Tjongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,
 Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 15 Row: 9 Column: 18
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 10434050.

FEATURES

Location/Qualifiers

1..2029

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:10280 IMAGE:3955573"

/tissue type="Skin, melanotic melanoma"

/clone_lib="NIH-MGC-20"

/lab_host="DH108-R"

/note="Vector: pOTB7"

67..1971

/codon_start=1

/product="Unknown (protein for MGC:10280)"

/protein_id="AAH02923.1"

/db_xref="GI:12804139"

/translation="MSLVPAIYIYPLNQLKGGIVNYGVVKKFKPPYLSKGDYD
 SVYIVDQNTVLTCLFSGNYEALPIYKNGDIVRHLKIOVKKETOITSSGFA
 SITFETGLGAPIPTSSRYNFTEDHKMVEALRWASTHSPSWTLKLCDDVPMO
 YDLCOLLGKAEVDGAFLLKAWDGTPTPPSRVLLDLVLEGLDLSHRLQNTI
 DILVDNHVARSLLKVSFLRIYLSLHFKLOSNSENOTMLSEFLHGGTGYGRIR
 VLPESNDVDQLKLESANLTANSDVICOSEDDSPSSGSVSLYEVECOQLSA
 TILDHQYLRPLCALKQAPQYRIRAKRSYKPRLFQSVKHLCPKCHLQLEVP
 HEGDLDIFODGATPTDYKQNTSLYDSKIWTRNOKRKVAVHFVKNNGILPLSNE
 CLLLEGLTSLFICKLSNKFNSVIVPSRSHEDLELLDLSAPFLIGTTHYCKDCSS
 LRSIONLSVDTKSWIPSSVAEALGIVPLOYVFTFLDDGTCVLEAYLMDSKFF
 QLPASFLMDLDLQKSDMIMDMFCPPGKIDAYPWLECFIKSYNVTNGTDNQICIQI
 FDTTVAEDVI"

655 a 391 c 386 g 597 t

BASE COUNT
ORIGIN

Query Match 80.7%; Score 1048; DB 9; Length 2029;
 Best Local Similarity 90.8%; Pred. No. 7.5e-268;
 Matches 1178; Conservative 0; Mismatches 0; Indels 120; Gaps 1;
 QY 1 atgtcttgggtccagcaaaattatatatatacacccctgaatcaacttaagggtggt 60
 DB 67 ATGCTTTGGTTCACGCAAAATATATATATACACCCCTGAATCAACTTAAGGTG 126
 QY 61 acaattgcaatgtctataggtgtgtgaagttctttaagcccccatatctaagcaagga 120
 DB 127 ACAATTGTCAATGCTATGGTGTGTGAAGTTCCTTTAAGCCCCCATATATTAAGCAAGGA 186
 QY 121 actgattattgtcagttgaattgttgagcagacaaatgtataaacttaacttgcctg 180
 DB 187 ACTGATTATTGCTCAGTGTGAATTTGTGGACGACAAATGTAAATTAACCTTGCCTTG 246
 QY 181 ctcttttagtggaaactatgaagcccttccaataattataaaatggagatattgttcgc 240
 DB 247 CTCTTTAGTGGAAACTATGAAGCCCTTCCAATAAATTTATAAAATGGAGATATTGTTC 306
 QY 241 ttccacaggctgaagattcaagtataaaaggagactcagggtatcaccagctctggc 300

RESULT 2

DB 307 TTTCACAGGCTGAAGATTCAAGTATATAAAAGGAGACTCAGGGTATACACCACTCTGGC 366
 QY 301 ttgcatctttgacgttttgaggaactttgggagccctatatactacgtcagcattcaagc 360
 DB 367 TTTGCACTCTTTGACGTTTGGAGGAACCTTTGGGAGCCCTATCATACCTCGCATCTCAAGC 426
 QY 361 aagtattttaacttctactactgaggaccacaaatggtagaagccttactggtttgtggca 420
 DB 427 AAGTATTTTAACTTCACTACTGAGGACCACAAATGGTAGAAGCCTTACGTGTTGGGCA 486
 QY 421 tctactcataatgacccgtcttgagcattactactaaaattgtgtgatgttcagcgaatgcag 480
 DB 487 TCTACTCATATGTACCGCTCTTGACATTACTATAAATTTGTGTATGTTTTCAGCAATGCAG 546
 QY 481 tatttgacctgactgtcagctcttgggcaagcagaagtgagcagagcattctttt 540
 DB 547 TATTTGACCTGACTTGTCACTCTTGGGCAAGCAGAGTGGACGGAGCATCATTTCTT 606
 QY 541 ctaaggatggatggcaccaggacacatttcccatcttggagagtccttaataacaagac 600
 DB 607 CTAAGGTATGGATGGCACCAGCACACCATTTCCCATCTTGGAGAGTCTTAATACAGAC 666
 QY 601 ctgttcttgaagtgatttaagtcacatccatcggtcacaaaatctgcacaatagacatt 660
 DB 667 CTGTGTTCTTGAAGGTGATTTAAGTCACATCCATCGGCTACAAAATCTGCAATAGACAT 726
 QY 661 ttagtctacgataaacatgttcatgtggcaagatctctgaagttggaagcttcttaaga 720
 DB 727 TTAGTCTACGATAAACCATGTTTCATGTGGCAAGATCTGGAAGTTGGAAGCTTCTTACA 786
 QY 721 atctatagcttcatcaccaaaacttcaatcaatgaattcagagagatcagacaatgtttaagt 780
 DB 787 ATCTATAGGCTTTCATACCAAACTTCAATCAATGAATTCAGAGAATCAGACAATGTTAAGT 846
 QY 781 tttaggttttcatcttcatgaggtaccagttacggtcggggaatcaggtcttgcagaa 840
 DB 847 TTAGAGTTTCACTTCATCGGAGGTACAGTTACGGTCGGGGAATCAGGCTCTTGCAGAA 906
 QY 841 agtaactctgagtgtggtcaactgaaagagatttagaactctgaaaatttgacagcgaat 900
 DB 907 AGTAACTCTGATGTGATCAACTCACTGAAAAGAGTTTAGAATCTGCAAAATTTGACAGCCAAT 966
 QY 901 cagcattcagatgttatctgtcaatcagaacctgcagcagagctttccaaaatggagctcgcg 960
 DB 967 CAGCATTTCAATGTTATCTGTCAATCAGAACCTGACGACAGCTTTTCCAA----- 1015
 QY 961 ctctgctccctcaggtcggagttcagtggcaggtctctcggtcatctgcagctccacctcc 1020
 DB 1016 ----- 1015
 QY 1021 tgagttcaagctctctctgctcctcagcctcccaagtagctgggattacagggtctggatca 1080
 DB 1016 -----GCTCTGGATCA 1026
 QY 1081 gtatcattatcagaggtagaagaagtcaacagctatctgtacaatacttacaagatcat 1140
 DB 1027 GTATCATTATACAGGTAGAAAGATGTCAACAGCTATCTGCTACAAATCTTACAGATCAT 1086
 QY 1141 cagattttggagaggacaccactatgtgccattttgaaacaaaagctcctcaacaatcac 1200
 DB 1087 CAGTATTTGGAGAGGACACCACCTATGTGCCATTTTGAACAAAAGCTCTCAACAATAC 1146
 QY 1201 cgcattccgagcaaaattgaggtcatataagccccagagaactatttcagctctgttaaatc 1260
 DB 1147 CCATCCGAGCAAAATGAGGTATATTAAGCCCAAGACTATTTCAGTGTGTTAACTT 1206
 QY 1261 cattgccctaaatgtcatttgcgcaagaagtctccaca 1298
 DB 1207 CATTCGCCCTAAATGTCATTGCTGCAAGAGTTCACACA 1244

SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 8.0%; Score 104.4; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 2.4e-21;
Matches 6; Conservative 468; Mismatches 104; Indels 0.

 \ddot{O} [illegible]

RESULT 14

AAF58262/c
ID AAF58262 standard; DNA: 936 BP.

AC AAF58262:

DT 24-APR-2001 (first entry)

XX DE Oligonucleotide D2007.

AA
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss. ,
KW

OS Synthetic.

PN WO200107665-A2.

AA PD 01-FEB-2001.

26-JUL-2000; 2000WO-US20476

XX 26-JUL-1999: 99US-0145695

PR 17-MAR-2000; 2000US-0190259.
XX

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

WPI: 2001-159728/16.

XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -

PS Example 6; Page 128; 159pp; English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match	8.0%	Score 104.4;	DB 22;	Length 936;
Best Local Similarity	0.8%	Pred. NO. 2.4e-21;		
Matches	6;	Conservative 468;	Mismatches 304;	Indels 0;
			Gaps	0;

[illegible]

CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX SQ Sequence 866 BP; 254 A; 162 C; 171 G; 274 T; 5 other;

Query Match 26.3%; Score 342; DB 22; Length 866;
Best Local Similarity 91.2%; Pred. No. 3.9e-93;
Matches 384; Conservative 0; Mismatches 30; Indels 7; Gaps 2;

OY 1 atgtcttgggtccagcaacaattatatatatatatacacccctgaacttaaggtggt 60
Db 439 atgtcttgggtccagcaacaattatatatatatatacacccctgaacttaaggtggt 498
OY 61 acaattgtcaatgtctatgtgtgtgaagttctttaagcccccatactcaagcaagga 120
Db 499 acaattgtcaatgtctatgtgtgtgaagttctttaagcccccatactcaagcaagga 558
OY 121 actgattattgtcagttgaactattgtggcagacacaaatgtataaactaacttgcctg 180
Db 559 actgattattgtcagttgaactattgtggcagacacaaatgtataaactaacttgcctg 618
OY 181 ctcttaagggaacatgaagcccttcacataattataaaatggagattgttcgc 240
Db 619 ctcttaagggaacatgaagcccttcacataattataaaatggagattgttcgc 678
OY 241 ttccacaggtgaagattcaagtataaaaggagacacaggtatcacccagctctg 300
Db 679 ttccacaggtgaagattcaagtntaaaggagacacaggtatcacccagctctg 738
OY 301 ttgcactttgacgtttgagggaactttggagagcccttatcatcactgcactcaagc 360
Db 739 ttgcactttgacgtt--gagggaactttggagagcccttatcatcactgcactnaagcc 796
OY 361 aagtatttaacttcactactgaggaccacacaaatgttagaagcctttagtggtggca 420
Db 797 agtatttacttact-----gagccncaaatgnaagccttacgggttgggct 851
OY 421 t 421
Db 852 t 852

RESULT 8
AAI61067
ID AAI61067 standard; cDNA; 2778 BP.
AC AAI61067;
DT 22-OCT-2001 (first entry)
DE Human polynucleotide SEQ ID NO 5056.
XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW anyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; SS.
XX Homo sapiens.
OS
XX WO200153312-A1.
PN

XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000WO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-052317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR P-PSDB; AAM41911.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Claim 1; SEQ ID NO 5056; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAH42213) with nontropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX SQ Sequence 2778 BP; 876 A; 495 C; 487 G; 920 T; 0 other;

Query Match 15.3%; Score 199; DB 22; Length 2778;
Best Local Similarity 100.0%; Pred. No. 1.1e-49;
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1100 aaagatgtcaacagctatctgtacataactacagatcatcagttatggagagacac 1159
Db 1 aaagatgtcaacagctatctgtacataactacagatcatcagttatggagagacac 60
OY 1160 cactatgtccattttgaaacaaaagctccctcaacataccgcatccgacaaaattga 1219
Db 61 cactatgtccattttgaaacaaaagctccctcaacataccgcatccgacaaaattga 120
OY 1220 ggtcatataagcccgagaagactatttcagttctgttaaaacttcctccataatgctatt 1279
Db 121 ggtcatataagcccgagaagactatttcagttctgttaaaacttcctccataatgctatt 180
OY 1280 tgcgtcaagaagttccaca 1298
Db 181 tgcgtcaagaagttccaca 199

RESULT 9
AAC28094
ID AAC28094 standard; cDNA; 336 BP.
XX

Db 1080 -----gctctggatca 1090
 QY 1081 gtatcattacagagtagaagatgtcaacagctatctgtctacatacttacagatcat 1140
 Db 1091 gtatcattacagagtagaagatgtcaacagctatctgtctacatacttacagatcat 1150
 QY 1141 cagttattgagagacaccactatgtgccattttgaaacaaaagctcctcaacaatc 1200
 Db 1151 cagttattgagagacaccactatgtgccattttgaaacaaaagctcctcaacaatc 1210
 QY 1201 cgcctcagcaaaattgaggtcattataagccagagactatttcagctctttaaactt 1260
 Db 1211 cgcctcagcaaaattgaggtcattataagccagagactatttcagctctttaaactt 1270
 QY 1261 cattgcctaaatgtcatttgcgtgcaagaagtccaca 1298
 Db 1271 cattgcctaaatgtcatttgcgtgcaagaagtccaca 1308

RESULT 5

AAH05303
 ID AAH05303 standard; cDNA; 576 BP.
 AC AAH05303;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (5'-primer) SEQ ID NO:2138.
 XX
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KW
 XX
 OS Homo sapiens.
 XX
 PW EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 98JP-0248036.
 XX
 PR 27-AUG-1999; 99JP-0300253.
 XX
 PR 11-JAN-2000; 2000JP-0118776.
 XX
 PR 02-MAY-2000; 2000JP-0183767.
 XX
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 1; SEQ ID 2138; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AA992446 to AA995893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX Sequence 576 BP; 167 A; 113 C; 114 G; 173 T; 9 other;

Query Match 42.0%; Score 545; DB 22; Length 576;

Best Local Similarity 98.7%; Pred. No. 1.5e-154;

Matches 545; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 atgtcttgggtccagcaacaattatatatatatacacccctgaatcaactaaaggtggt 60

Db 25 atgtcttgggtccagcaacaattatatatatatacacccctgaatcaactaaaggtggt 84

QY 61 acaattgtcaatgtctatgtgtgtgaagtcttttaagcccccatatcttaagcaaaagga 120

Db 85 acaattgtcaatgtctatgtgtgtgaagtcttttaagcccccatatcttaagcaaaagga 144

QY 121 actgattatgtcagttgttaactattgtgaccagacaaatgtataaactaacttgctg 180

Db 145 actgattatgtcagttgttaactattgtgaccagacaaatgtataaactaacttgctg 204

QY 181 ctcttagtgaaactatgaagcccttccataattataaaaatgagagatatgttcgc 240

Db 205 ctcttagtgaaactatgaagcccttccataattataaaaatgagagatatgttcgc 264

QY 241 ttccacaggtgaaagattcaagtataaaaaggagactcagggtatcaccagctctggc 300

Db 265 ttccacaggtgaaagattcaagtataaaaaggagactcagggtatcaccagctctggc 324

QY 301 ttgcatctttgacgtttgagggaaactttggagccctcatcacctgcacttcaagc 360

Db 325 ttgcatctttgacgtttgagggaaactttggagccctcatcacctgcacttcaagc 384

QY 361 aagtattttaaacttcactactgaggaccacacaaatgtagaagccttactgtttgggca 420

Db 385 aagtattttaaacttcactactgaggaccacacaaatgtagaagccttactgtttgggca 444

QY 421 tctactcatatgtcacctgttggacattactataaaattgtgtatgttcagccaatgcag 480

Db 445 tctactcatatgtcacctgttggacattactataaaattgtgtatgttcagccaatgcag 504

QY 481 tatttgacctgacttgcagctcttggcacaagcagaagtgagcagagcatcttctt 540

Db 505 tatttgacctgacttgcagctcttggcacaagcagaagtgagcagagcatcttctt 564

QY 541 cttaaaggtatgg 552

Db 565 cttaaaggtatgg 576

RESULT 6

AAH07649
 ID AAH07649 standard; cDNA; 854 BP.
 AC AAH07649;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (5'-primer) SEQ ID NO:4484.
 XX
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KW
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.

DE Human cDNA sequence SEQ ID NO:11187.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

KW Homo sapiens.

OS EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Claim 8; SEQ ID 11187; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX Sequence 2631 BP; 826 A; 473 C; 491 G; 841 T; 0 other;

Query Match 80.7%; Score 1048; DB 22; Length 2631;
Best Local Similarity 90.8%; Pred. No. 3.6e-306;
Matches 1178; Conservative 0; Mismatches 0; Indels 120; Gaps 1;

Qy 1 atgtcttgggtccagcaacaattatataacacccctgaatcaacttaagggtggt 60
Db 24 atgtcttgggtccagcaacaattatataacacccctgaatcaacttaagggtggt 83
Qy 61 acaattgtcaatgtctatgtgtgtgaagttctttaagcccccatatctaaagcaagga 120
Db 84 acaattgtcaatgtctatgtgtgtgaagttctttaagcccccatatctaaagcaagga 143
Qy 121 actgattattgtcagttgaactattgtgagcagacaataatgtaaacctgctgt 180
Db 144 actgattattgtcagttgaactattgtgagcagacaataatgtaaacctgctgt 203

Qy 181 ctctttagtgaactatgaagcccttccaataattataaaatggagatatgttcgc 240
Db 204 ctctttagtgaactatgaagcccttccaataattataaaatggagatatgttcgc 263
Qy 241 ttccacaggtgaagattcaagtataaaaagagactcaggttatcacagctctggc 300
Db 264 ttccacaggtgaagattcaagtataaaaagagactcaggttatcacagctctggc 323
Qy 301 ttgcatctttgacgtttgagggaactttggagccctatcatacctcgcaactcaagc 360
Db 324 ttgcatctttgacgtttgagggaactttggagccctatcatacctcgcaactcaagc 383
Qy 361 aagattttaacttcaactactgagacacacaaaatggtagaagccttacgtgttgggca 420
Db 384 aagattttaacttcaactactgagacacacaaaatggtagaagccttacgtgttgggca 443
Qy 421 tctactcatatgcaccgtcttgggacattactaaaattgtgtgatttcagccaatgcag 480
Db 444 tctactcatatgcaccgtcttgggacattactaaaattgtgtgatttcagccaatgcag 503
Qy 481 tatttgacctgactgttcagctcttgggcaaaagcagaagtgagcggagcatcttctt 540
Db 504 tatttgacctgactgttcagctcttgggcaaaagcagaagtgagcggagcatcttctt 563
Qy 541 cttaaagggtatgggtggtgacccaggacacatttccatcttggagagtcttaatacaagc 600
Db 564 cttaaagggtatgggtggtgacccaggacacatttccatcttggagagtcttaatacaagc 623
Qy 601 ctgttcttgaagggtgatttaagtcacatccatcggtacacaaaatcgcaaatagacatt 660
Db 624 ctgttcttgaagggtgatttaagtcacatccatcggtacacaaaatcgcaaatagacatt 683
Qy 661 tttagtctacataaacatgttcatgtggcaagatctctgaaggttggaggtcttcttaga 720
Db 684 tttagtctacataaacatgttcatgtggcaagatctctgaaggttggaggtcttcttaga 743
Qy 721 attctagcttcataccaaaattcaaatcaatgaattcagagaatcagacaattgtaagt 780
Db 744 attctagcttcataccaaaattcaaatcaatgaattcagagaatcagacaattgtaagt 803
Qy 781 tttagatttctcatctcatggaggtaccagttacggttcgggggaatcaggtgtcttcagaa 840
Db 804 tttagatttctcatctcatggaggtaccagttacggttcgggggaatcaggtgtcttcagaa 863
Qy 841 agtaactctgattggtgatacaactgaaaaggatttagaattcgcaaatgtgacagccaat 900
Db 864 agtaactctgattggtgatacaactgaaaaggatttagaattcgcaaatgtgacagccaat 923
Qy 901 cagcattcagatgttatctgtcaatcagaaacctgacgacagctttccaaatggagctcg 960
Db 924 cagcattcagatgttatctgtcaatcagaaacctgacgacagctttccaaatggagctcg 972
Qy 961 ctctgtctctccagggtggtgaggttcagtggtgacggtctcggtcattgacgctccacctcc 1020
Db 973 ctctgtctctccagggtggtgaggttcagtggtgacggtctcggtcattgacgctccacctcc 972
Qy 1021 tgagttcaagcttctctgctcagcctcccaagtagctgggattacaggtcttgatca 1080
Db 973 tgagttcaagcttctctgctcagcctcccaagtagctgggattacaggtcttgatca 983
Qy 1081 gtatcattatacagagtagaagaatgtcaacagctatctgctacaatacttacagatcat 1140
Db 984 gtatcattatacagagtagaagaatgtcaacagctatctgctacaatacttacagatcat 1043
Qy 1141 caqtatttgagagagacacacactatgtccattttgaaacaaaagctctccaacatc 1200
Db 1044 caqtatttgagagagacacacactatgtccattttgaaacaaaagctctccaacatc 1103
Qy 1201 cgcctccgagcaaaaattgaggttcataaagcccgagaagactatttcagctgtgttaactt 1260
Db 1104 cgcctccgagcaaaaattgaggttcataaagcccgagaagactatttcagctgtgttaactt 1163

full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8: SEQ ID 12760; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2077 BP; 657 A; 401 C; 393 G; 626 T; 0 other;

Query Match 80.7%; Score 1048; DB 22; Length 2077;
Best Local Similarity 90.8%; Pred. No. 3.2e-306;
Matches 1178; Conservative 0; Mismatches 0; Indels 120; Gaps 1;

1 atgtcttgggtccagcaaaaattatataatataccacctgaatcaactaaagggtgt 60
|||||
131 atgtcttgggtccagcaaaaattatataatataccacctgaatcaactaaagggtgt 190

61 acaatgtcaatgtcataggtgtgtgaagttctttaagccccccatataatgaagca 120
|||||
191 acaatgtcaatgtcataggtgtgtgaagttctttaagccccccatataatgaagca 250
|||||
121 actgattattgtcagttgaactatttggaccagacaaatgtaaaactaacttgcctg 180
|||||
251 actgattattgtcagttgaactatttggaccagacaaatgtaaaactaacttgcctg 310
|||||
181 ctcttagtgaactatgaagcccttccaaataattataaaaatggagatatgttcgc 240
|||||
311 ctcttagtgaactatgaagcccttccaaataattataaaaatggagatatgttcgc 370
|||||
241 ttccacaggtcgaagattcaagtataaaaaggagactcagggtatcacccagctctgc 300
|||||
371 ttccacaggtcgaagattcaagtataaaaaggagactcagggtatcacccagctctgc 430
|||||
301 ttgcatctttgacgtttgagggaactttggagccccctatcatacctcgcaactcaagc 360
|||||
431 ttgcatctttgacgtttgagggaactttggagccccctatcatacctcgcaactcaagc 490
|||||
361 aagatttttaacttcaactactagagaccacaaatgttagaagccttactgtttgggca 420
|||||
491 aagatttttaacttcaactactagagaccacaaatgttagaagccttactgtttgggca 550
|||||
421 tctactcatatgtcacccgtctggacattactaaaaattgtgtgatttccagccaatgcag 480
|||||
551 tctactcatatgtcacccgtctggacattactaaaaattgtgtgatttccagccaatgcag 610
|||||
481 tattttgacctgactgtcagctcttggcacaagagaagtggagcagcagcattcttt 540
|||||
611 tattttgacctgactgtcagctcttggcacaagagaagtggagcagcagcattcttt 670
|||||
541 ctaaaaggtatggatggcaccaggacacatttccattcttggagagtcttaatacaagac 600

Db 671 ctcaagggtatggatggcaccaggacaccatttccattcttgagagttcttaatacaagac 730
QY 601 cttgtcttgaagggtgatttaagttcacatccatcggtctacaaaaatctgacaatagacatt 660
Db 731 cttgtcttgaagggtgatttaagttcacatccatcggtctacaaaaatctgacaatagacatt 790
QY 661 ttatgtctacgataaacatgttctatgtgccaagatctctgaaggttggagtttttttaga 720
Db 791 ttatgtctacgataaacatgttctatgtgccaagatctctgaaggttggagtttttttaga 850
QY 721 actatagccttcataccaaaacttcaatcaatgaattcagagaatcagacaatgttaagt 780
Db 851 actatagccttcataccaaaacttcaatcaatgaattcagagaatcagacaatgttaagt 910
QY 781 ttatgagtttcatcttcacggaggtaccaggttacggttcggtggaatcaggttcttgcagaa 840
Db 911 ttatgagtttcatcttcacggaggtaccaggttacggttcggtggaatcaggttcttgcagaa 970
QY 841 agtaactctgatgtggatacaactgaaaaggatttagaattcgaatttgcagagccaat 900
Db 971 agtaactctgatgtggatacaactgaaaaggatttagaattcgaatttgcagagccaat 1030
QY 901 cagcattcagatgttatctgtcaatcagaacctgcagacagctttccaaatggagttctcg 960
Db 1031 cagcattcagatgttatctgtcaatcagaacctgcagacagctttccaaatggagttctcg 1079
QY 961 ctctctctccaggctggaggttcagtggaacggtctcggtctcatttgcagctccacctcc 1020
Db 1080 ----- 1079
QY 1021 tgagttcaagcttctctcgtcctcagcctcccaagtagctgggattacagggctctggatca 1080
Db 1080 -----gctctggatca 1090
QY 1081 gtatcattatcacgaggtgaaagatgtcaaacagctatctgtcacaaatcttacagatcat 1140
Db 1091 gtatcattatcacgaggtgaaagatgtcaaacagctatctgtcacaaatcttacagatcat 1150
QY 1141 cagatttggagagagacaccactatgtccattttgaaacaaaagctccccaacaatc 1200
Db 1151 cagatttggagagagacaccactatgtccattttgaaacaaaagctccccaacaatc 1210
QY 1201 cgcattccagcaaaattgaggttcataatgaagccagagaactatttcagctgttcaactt 1260
Db 1211 cgcattccagcaaaattgaggttcataatgaagccagagaactatttcagctgttcaactt 1270
QY 1261 cattgcccataatgtcatttgcgtcgaagaagtccaca 1298
Db 1271 cattgcccataatgtcatttgcgtcgaagaagtccaca 1308

RESULT 2

AAH17603

ID AAH17603 standard; cDNA: 2383 BP.

XX AC AAH17603;

XX XX 26-JUN-2001 (first entry)

XX XX Human cDNA sequence SEQ ID NO:17112.

XX DE Human: primer: detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX XX 07-FEB-2001.

XX XX 28-JUL-2000; 2000EP-0116126.

XX XX 29-JUL-1999; 95JP-0248036.

```

Oy 1126 atactcacagatcatgatttggagagacacaccactatgtgccatttttgaacaaaaa 1185
Db 242 ATACTTACAGATCAATCAGTATTGGAGAGGACACCACTATGTGCCATTTTGAACAACAAA 301

Oy 1186 gctcctcaacaatacccgatccgagcaaaattgaggtcatataagcccgagaagactattt 1245
Db 302 GCTCCTCAACAATACCGCATCCGAGCAAAATTGAGGTCTATATAAGCCCGAAGACATTT 361

Oy 1246 cagctctgtaaaacttcattcccttaaatgtca-ttctgtcgaagaagttccaca 1298
Db 362 CAGTCTGTAAACTTTCATTCGCCCTAAATGTCATTTTGTGCTCAAGAAGTTCCACA 415

RESULT 15
AV656827 328 bp mRNA EST 07-SEP-2000
LOCUS AV656827 GLC Homo sapiens cDNA clone GLCEVD06 3', mRNA sequence.
DEFINITION AV656827
ACCESSION AV656827
VERSION AV656827.1 GI:9877841
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 328)
Xian,B., Wu,I., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H.,
, Xu,X., Li,N., Peng,Y., Liu,F., Ou,J., Song,H., Cheng,Z., Ou,J.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
, G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
Homo sapiens cDNA clone
Unpublished (2000)
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanqgchgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..328
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCEVD06"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 103 a 58 c 55 g 111 t 1 others
ORIGIN

Query Match 20.6%; Score 268; DB 10; Length 328;
Best Local Similarity 100.0%; Pred. No. 9.4e-58;
Matches 268; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 atgtcttgggtccagcaacaattatatatatacacccctgaatcaacttaagggtggt 60
Db 45 ATGCTTTGTTGCCAGCAACAATATATATATACACCCCTGAATCAACTTAAAGGGTGGT 104

Oy 61 acaattgtcaatgtctatgtgtgtgtggaagttctttaagcccccatcttaagcaagga 120
Db 105 ACAATTGTCAATGCTATGGTGTGTGGAAGTCTTTAAGCCCCCATATCTAAGCAAGGA 164

Oy 121 actgattattgtcagttgtaactattgtgaccgagacaaatgtaaaactaactgcctg 180
Db 165 ACTGATTATTGCTCAGTTGTAAGTATTGTGGACACAGACAAATGTAAAACTAACTTGCCTG 224

Oy 181 ctctttagtggaaactatgaagcccttccaataattataaaaatggagataattgttcgc 240
Db 225 CTCTTTAGTGGAAACTATGAAGCCCTTCCCAATAATTATATAAAATGGAGATATTGTCGC 284

```

```

Oy 241 ttccacagcgtgaagattcaagtatafa 268
Db 285 TTTCACAGGCTCAAGATTCAAGTATATA 312

```

Search completed: April 9, 2002, 22:13:58
Job time: 8873 sec

pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcggg
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH-MGC Library.*

BASE COUNT 302 a 197 c 205 g 299 t 1 others
ORIGIN

Query Match 24.4%; Score 317.2; DB 11; Length 1004;
Best Local Similarity 83.6%; Pred. No. 3.5e-70;
Matches 443; Conservative 0; Mismatches 78; Indels 9; Gaps 7;

Oy 1 atgtcttgggtccagcaacaattatatatacacccctggaactcaacttaagggtggt 60
|||||
Db 432 ATGCTTTGGTTCAGCAACAANAATATATATACACCCCTGAATCAACTTAAGGGIGGT 491
|||||
Oy 61 acaattgtcaatgtctatgtgtgtgaagttctttaagcccccatatctaaagcaagga 120
|||||
Db 492 ACAATGTCAATGTCTATGGTGTGTGAAGTCTTTAAAGCCCCCATATCTAAGCAAGGA 551
|||||
Oy 121 actgattatgtcagttgaactatttggaccagacaaaatgtaaaactaacttgcctg 180
|||||
Db 552 ACTGATTATGCTCAGTGTGAATTTGGACGACGACAAATGT-AACTAACITTCCTG 610
|||||
Oy 181 ctcttttagtgaactatgaagcccttccaataattata--aaatggagatatgttctc 238
|||||
Db 611 CTCTTTAGTGAACATATGAAGCCCTTCQAATTAATTAACACAAATGGAGAATTTGTT 670
|||||
Oy 239 gctttcacaggtggaattcaagtatataaaaggagagactcaggtatcaccagctctg 298
|||||
Db 671 GCTTTACAGGCTGAGATTCAGATATATACAGGAGACTCA-GGTATCACCAGTCTG 729
|||||
Oy 299 gctttgcatctttgacgtttgagggaactttggagccctatcatcaccctgcacttcaa 358
|||||
Db 730 GCTTTGCATC-TTGAGCTCTGAAGAACTTGGAGCCCTATCATATACTACGCGCATTC 788
|||||
Oy 359 gcaagtatttaacttactactagga--ccacaaaatggttagaa--gccttactgttt 415
|||||
Db 789 GCCAGTATTTAACCTTCCTTACTGTAGGACCCACAAATGTTAGAAGGCCCTTCGTCTGTG 848
|||||
Oy 416 gggcatctactatgtcaccgctcttggacattactaaattgtgtgattgttcagccaa 475
|||||
Db 849 GGGCATCTACCCATATGTTACCGGCTCTGCAATTTCTCACA-TGTGTGGAGTTCAGCACA 907
|||||
Oy 476 tgcagtatttgcactgaacttgcagctcttgggcaagacagaagtggac 525
|||||
Db 908 GGACGTATAGAAACGAGATTGTGCCCCGTGCAAGAAATGTCAGAGGAC 957
|||||

RESULT 12
AA382297 341 bp mRNA EST 21-APR-1997
LOCUS EST95690 Testis 1 Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION
ACCESSION AA382297
VERSION AA382297.1 GI:2034796
KEYWORDS
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 341)
REFERENCE
ADAMS, M.D., SOARES, M.B., KERLAVAGE, A.R., FIELDS, C. and VENTER, J.C.
Rapid cDNA sequencing (expressed sequence tags) from a
directionally cloned human infant brain cDNA library;
Nature Genet. 4, 373-380 (1993)
JOURNAL
MEDLINE 94004965
Other ESTs: THC121679
Contact: Kerlavage, AR
Bioinformatics

REFERENCE
ADAMS, M.D., SOARES, M.B., KERLAVAGE, A.R., FIELDS, C. and VENTER, J.C.
Rapid cDNA sequencing (expressed sequence tags) from a
directionally cloned human infant brain cDNA library;
Nature Genet. 4, 373-380 (1993)
JOURNAL
MEDLINE 94004965
Other ESTs: THC121679
Contact: Kerlavage, AR
Bioinformatics

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavet@tigr.org

For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES
Location/Qualifiers
source 1..341

/organism="Homo sapiens"
/db_xref="ATCC (inhost):186770"
/db_xref="taxon:9606"
/clone_lib="Testis 1"
/sex="male"
/dev_stage="adult"

/note="Organ: testis; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI"
BASE COUNT 108 a 68 c 67 g 98 t
ORIGIN

Query Match 24.0%; Score 311.8; DB 10; Length 341;
Best Local Similarity 99.1%; Pred. No. 6.8e-69;
Matches 324; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Oy 625 cacatccatcaggtacaaaatcgcacaaatcttagtctacgataaccatgttcat 684
|||||
Db 1 CACATCCATCGGCTACAAATCTGACATATGACATTTTAGTCTAGATAACCATGTTCA 60
|||||
Oy 685 gtggcaagatctctgaaggttggaagctttcttagaatctatagccttcataccaaactt 744
|||||
Db 61 GTGGCAAGATCTCTCAAGGTTGGAAGCTTCTTAGAATCTATAGCCTTCATACCAACTT 120
|||||
Oy 745 caatcaatgaattcagagaatcagacaatgttaagttagatttcattcattcagaggt 804
|||||
Db 121 CAATCAATGAATTCAGAAATCAGACAAATGTTAAGTTTAGAGTTTCACTTCATGGAGGT 180
|||||
Oy 805 accagttacggtcgggggaatcagggtcttggcagaaagtaacctgtatgttgatacaactg 864
|||||
Db 181 ACCAGTTACGCTCGGGGAATCAGGCTCTGCCAGAAAGTAACCTGTGATCAACTG 240
|||||
Oy 865 aaaaaggatttagaactctgaaaatttgacaccaa--tcagcattcagatgttatctgca 923
|||||
Db 241 AAAAAGGATTTAGAATCTGCAAAATTTGACAGCAATTCAGCAATTCAGATGTATCTGTCA 300
|||||
Oy 924 atcagaacctgacgacagctttccaaa 950
|||||
Db 301 ATCAAGAACTGACGACAGCTTTTCCAA 327
|||||

RESULT 13
BG088861 565 bp mRNA EST 26-JAN-2001
LOCUS BG088861
DEFINITION H3158E09-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
ACCESSION BG088861
VERSION BG088861.1 GI:12571425
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 565)
REFERENCE
KARGUL, G.J., DUDKULA, D.B., QIAN, Y., LIM, M.K., JARADAT, S.A., TANAKA
T.S., CARTER, M.G. and KO, M.S.H.
Verification and initial annotation of NIA mouse 15K cDNA clone set
Unpublished (2001)
Other ESTs: H3158E09-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health


```

Db 418 ATCAATGACATCCAGAGATCAGACAA-GTTAACGTCAGAGTCTCATCTTCATGGAGG 476
Qy 804 taccagtacgggtcgggaatcagggtcttgcagaaa 841
Db 477 TACCAGTTACGGTCGGGGCATCAGGCTTGGCCAAGAA 514

RESULT 8
AUI24189 AUI24189 866 bp mRNA EST 23-OCT-2000
LOCUS AUI24189 NT2RM2 Homo sapiens cDNA clone NT2RM2001805 5', mRNA
DEFINITION sequence.
ACCESSION AUI24189
VERSION AUI24189.1 GI:10948905
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 866)
AUTHORS Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
,Y., Sugano,S., Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1..866
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM2001805"
/clone_lib="NT2RM2"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"
BASE COUNT 254 a 162 c 171 g 274 t 5 others
ORIGIN
1 atgtcttgggtccagcaacaattatatatatataccaccctgaatcaacttaagggtggt 60
439 ATGCTTTGGTCCACACAAATATATATATACACCCCTGAATCAACTTAAGGGTGGI 498
Qy 61 acaattgtcaatgtctatggtgtgtggaagttctttaagcccccatatctaaagcaagga 120
Db 499 ACAATTGTCAATGCTCTATGGTGTGTGAAGTCTTTAAGCCCCCATATCTAAGCAAGGA 558
Qy 121 actgattattgtcagttgtaactattgtggaccgacaaaatgtaaaactaactgacctg 180
Db 559 ACTGATTATGCTCAGTTGTAAGTCTATTGTGGACCAACAAATGTAAACTTAACTTGCCTG 618
Qy 181 ctctttagtggaactcatgaagcccttccaataattataaaatggagatatgttcgc 240
Db 619 CTCTTTAGTGGAACTATGAAGCCCTTCCAATAATTATATAAAATGGAGATATTGCTGCC 678

```

```

Qy 241 tttcacagggtgaagattcaagtataataaaaggagactcagggtatcacacagctctggc 300
Db 679 TTTACAGGCTCAAGATTCAAGTNTAAAAAGAGACTCAGGGTATCACCAGCTCTGGC 738
Qy 301 tttgcattcttgaactttgagggaaactttggagccctctatataccctcgcaactcaagc 360
Db 739 TTTGCACTCTTTGACGTT--GAGGGACTTTGGGAGCCCCCTATCATACCTTGCACTTNA GCC 796
Qy 361 aagtatatttaacttcaactactagagcaccacaaaatgtagaagccttcagtggtttgggca 420
Db 797 AGTATTTTACTTCTCTACT-----GAGGCCNCAAAATGGNAGAAGCCTTACGGGTTTGGGCT 851
Qy 421 t 421
Db 852 t 852

RESULT 9
AUI24189 AUI24189 409 bp mRNA EST 15-OCT-1999
LOCUS AUI24189 NT2RM2 Homo sapiens cDNA clone
DEFINITION IMAGE:2581377 3', mRNA sequence.
ACCESSION AUI24189
VERSION AUI24189.1 GI:6043420
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 409)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Gibco
High quality sequence stop: 403.
FEATURES
source
1..409
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2581377"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP-CCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
1.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.
BASE COUNT 120 a 79 c 74 g 136 t
ORIGIN
1 tttcacagggtgaagattcaagtataataaaaggagactcagggtatcacacagctctggc 300
679 TTTACAGGCTCAAGATTCAAGTNTAAAAAGAGACTCAGGGTATCACCAGCTCTGGC 738
301 tttgcattcttgaactttgagggaaactttggagccctctatataccctcgcaactcaagc 360
739 TTTGCACTCTTTGACGTT--GAGGGACTTTGGGAGCCCCCTATCATACCTTGCACTTNA GCC 796
361 aagtatatttaacttcaactactagagcaccacaaaatgtagaagccttcagtggtttgggca 420
797 AGTATTTTACTTCTCTACT-----GAGGCCNCAAAATGGNAGAAGCCTTACGGGTTTGGGCT 851
421 t 421
852 t 852

Query Match 25.9%; Score 336.6; DB 10; Length 409;
Best Local Similarity 98.8%; Pred. No. 3.5e-75;
Matches 339; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 609 tgaagggtgatttaagtcacatccatcggtcacaaaatctcacaaatagacatttagtcta 668
Db 409 TGAAGGTGATTTAAGTCACATCCCATCGCTACAAAATCTGACAAATAGACATTTTAGTCTA 350
Qy 569 cgataaccatgttcatgtggcaagatctctgaaggttggagcttttcttagaatctatag 728

```

/clone="PLACE1005026"
/clone_lib="PLACE1"
/tissue_type="placenta"
/note="Vector: PME185FL3"

BASE COUNT 259 a 164 c 166 g 259 t 6 others
ORIGIN

Query Match 40.2%; Score 522.4; DB 10; Length 854;
Best Local Similarity 99.3%; Pred. No. 2.2e-122;
Matches 545; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 atgtcttgggtccagcaacaataatatatatatacccccgtgaatcaacttaagggtgtt 60
Db 132 ATGTCTTTGGTTCAGCAACAATAATATATATATACACCCCTGAATCAACTTAAGGGTGT 191
QY 61 acaattgcaatgtctatggtgtgtgaagttctttaagcccccatacttaagcaagga 120
Db 192 ACAATTTGCAATGTCTATGGTGTGTGAAGTTCCTTAAGCCCCCATATCTTAAGCAAGGA 251
QY 121 actgattattgtctcagttgtaactattgtggaccagacaaatgtaaaaacttaacttgcctg 180
Db 252 ACTGATTATTGCTCAGTTGTAACATTGTGGACACACAAATGTAAACTTAACCTTGCCCTG 311
QY 181 ctcttttagtggaaactatgaagcccttccaataattataaaaaatggagatatattgtcgc 240
Db 312 CTCTTTAGTGGAACTATGAAGCCCTTCCAATAATTTATAAAATGGAGATATTGTTCGC 371
QY 241 ttccagagctgaagattcaagtataataaaaggagactcagggtatcacccagctctggc 300
Db 372 TTTCACAGCTGAGATTCAGTATATAAAGAGGACATCAGGGTATACACAGCTCTGGC 431
QY 301 ttgtcatctttgacgtttgagggaaactttgggagcccttatcatcacctgcacttcaagc 360
Db 432 TTTCGATCTTTGACGCTTTGAGGGAACCTTTGGGAGCCCTTATCATACCTCGCATTCAGC 491
QY 361 aagatttttaacttcactactagacacacaaaatggtagaagcctttagctgtttgggca 420
Db 492 AGTATTTTAACTTCACTACTAGGACCCACAAAATGGTAGAGCCTTACGTTGTTGGGCA 551
QY 421 tctactcatatgtcccgcttcttggacattactataaattgtgtgtgttcagcc-aatgca 479
Db 552 TCTACTCATATGTCCCGCTCTTGGACATTACTTAAATTTGTGTGTGTTCAGGCCAAATGCA 611
QY 480 gtatttttaacttgcactgtcagctcttggcagaagcagaagtgagc-ggagcatcatttc 538
Db 612 GTATTTTACCTTGAATTTGCAACTCTTGGCAAGCAAGAGTGGAGGNGCATCAITTC 671
QY 539 ttctaaagg 547
Db 672 TTCTAAAGG 680

RESULT 5
AA351459/c 429 bp mRNA EST 21-APR-1997
LOCUS EST59213 Infant brain Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION AA351459
ACCESSION AA351459
VERSION AA351459.1 GI:2003779
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 429)
AUTHORS Adams,W.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
TITLE Rapid cDNA sequencing (expressed sequence tags) from a
directionally cloned human infant brain cDNA library
JOURNAL Nature Genet. 4, 373-380 (1993)
MEDLINE 94004965
COMMENT Other ESTs: EST59212 THC121679
Contact: Kerlavage, AR
Bioinformatics

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES
Source

1. 429
/organism="Homo sapiens"
/db_xref="ATCC (inhost):151767"
/db_xref="taxon:9606"
/clone_lib="Infant brain"
/sex="female"
/dev_stage="Infant"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI"

BASE COUNT 125 a 94 c 84 g 125 t 1 others
ORIGIN

Query Match 29.7%; Score 385.2; DB 10; Length 429;
Best Local Similarity 97.2%; Pred. No. 1.5e-87;
Matches 413; Conservative 0; Mismatches 9; Indels 3; Gaps 2;

QY 399 aagaagccttaccgtgttgggcatc--tactcataatgtcacctgttggac-attactaaa 455
Db 425 AGAAGCCTTACGTTGTTGGGCAACCTACTCCATATNCACCGCTCTGGGACAATTACTAAA 366
QY 456 attgtgtgattgttcagccaatcagattatttgaactgacttgcagctcttgggcaaacg 515
Db 365 ATTGTGTATGTTCAGCCCAATGCAGTATTTCACCTGACTTGTGACGCTCTTGGGCAAGC 306
QY 516 aagaagtgacggagcatcattcttctaaagtgatggatggcaccagacaccatttc 575
Db 305 AGAAGTGGACGGAGCATCATTTCTTAAAGGTATGGATGGCACCAGACACCATTTAC 246
QY 576 atcttggagagctcttaatacaagacccttcttgaagggtgatttaagtcacatccatcg 535
Db 245 ATCTTGGAGAGCTTTAATACAAAGACCTTGTCTTGAAGGTGATTAAAGTCACATCCATCG 186
QY 636 gctcaaaaatctgacaatagacatttttagtctacgataaccatgttcatgttgccaagatc 695
Db 185 GCTACAAATCTGACAATAGACATTTTAGTCTACGATAACCATGTTTCATGTGCAAGATC 126
QY 696 tctgaaggttggagccttctttagaatctatagccttcataccaaacttcaatcaatgaa 755
Db 125 TCTGAAGGTTGGAGCTTTCTTTAGAATCTATAGCCTTCATACCAACTTCAATCAATGAA 66
QY 756 ttcagagaatcagacaatgtttaagtttagagtttcatcttcatgaggtaccagttacgg 815
Db 65 TTCAGAGAAATCAGACAAATGTTAAGTTTAGAGTTTTCATCTTTCATGAGGTTACCGG 6

RESULT 6

AL520360
LOCUS AL520360 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB006YH17 5
DEFINITION prime, mRNA sequence.
ACCESSION AL520360
VERSION AL520360.1 GI:12783853
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 782)

/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 271 a 205 c 212 g 238 t
ORIGIN

Query Match 47.1%; Score 611.2; DB 11; Length 925;
Best Local Similarity 94.0%; Pred. No. 5.8e-145;
Matches 679; Conservative 0; Mismatches 36; Indels 5; Gaps 4;
QY 87 gaagttcttaagcccccatactcaagaaagaaactaattattgctcaagttgaactat 146
|||||
Db 1 GAAGTCTTTAGCCCCCAATCTATAGCAAGAACTGATTATTGCTCAGTTGTAACTAT 60
QY 147 tgtggaccagacaaatgtaaaactaaactgctgctctcttadtgtgaaactatgaagccct 206
|||||
Db 61 TGTGGACCAGACAAATGTAAACTAACTTTCCTGCTCTTTAGTGGAACTATGAAGCCCT 120
QY 207 tcaataattataaaatggagatattgttcgcttcaagagctgaagattcaagtata 266
|||||
Db 121 TCCAAATTAATTTAAATAATGGAGATATTGTCGCTTTACAGGCTGAAGATCAAGTATA 180
QY 267 taaaaggagactcaggggtatcacagctctgcttgcattcttgcagtttgaagaaac 326
|||||
Db 181 TAAAAGGAGACTCAGGGTATCACAGCTCTGCTTTGCACTTTTGCAGCTTTGAGGGAAAC 240
QY 327 ttgggagccctatacatacctcgcaactcaagcaagtaattttaacttcaactgaagga 386
|||||
Db 241 TTTGGAGCCCTATCATACCTCGCACTTCAAGCAAGTATTTAACTTCACTACTAGGA 300
QY 367 ccacaaatggtagaagccttgcgtttgtggcactactactataatgcagctctggac 446
|||||
Db 301 CCACAAATGGTAGAAGCCTTACGTTTGGGCATCTACTATATGTACCGCTCTGGAC 360
QY 447 attactaaaattgtgtgattgttcagccaatgcagtatatttgcacctgaactgtcagctctt 506
|||||
Db 361 ATTACTAAAATTTGTGTATGTTCCGCAATGCAGTATTTTGACCTGACTTGTCACTCTT 420
QY 507 gggcaagcagaagtggcagagacatcttcttaaggtatggatggcaccagac 566
|||||
Db 421 GGGCAAGCAGAAGTGGACGGAGCATCATTTCTTAAAGGTATGGGATGGCACCAGGAC 480
QY 567 accatttccatcttgagagctcttaacaagacctgtctctgaagtgatttaagtca 626
|||||
Db 481 ACCATTTCATCTGGAGAGCTTAATACAGACCTTGTCTTGAAGGTGATTTAAGTCA 540
QY 627 catcaatcggtacaaaatctgacaatagacatttttagtctacgataaaccatgttcatgt 686
|||||
Db 541 CATCCATCGG-TACAAAATCTGACAAATAGACATTTTAGTCTACGATAACCATGTTTCATGT 599
QY 687 g-gcaagatctcgaaggttggaagcttcttagaactctatagctctcataccacaaacttc 745
|||||
Db 600 GCGCCAGATCTCTGAAGTTGGAAG-ATTCCTAGAACTATATAGCTCTTAGCCAA--CTT 656
QY 746 aatcaatgaattcagagaatcagacaatgttaagtttagagtttcttcttcttcttcttcttctt 805
|||||
Db 657 CATCATGGAATTCGGAGAAATCGGACACGGTTTCAGTTTTCAGGTCATCTCCAGGAGGCCA 716
QY 806 CC 807
Db 717 CC 718

RESULT 2
BF027455
LOCUS BF027455 920 bp mRNA EST 10-OCT-2000
DEFINITION 601672751F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3955573 5',
mRNA sequence.
ACCESSION BF027455
VERSION BF027455.1 GI:10735167
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 920)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCMB30 row: e column: 14
High quality sequence start: 52
High quality sequence stop: 746.

FEATURES
source

Location/Qualifiers
1..920
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3955573"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
ECORI; CDNA made by oligo-dr priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 275 a 189 c 189 g 267 t
ORIGIN

Query Match 42.3%; Score 548.6; DB 11; Length 920;
Best Local Similarity 92.0%; Pred. No. 4.9e-129;
Matches 702; Conservative 0; Mismatches 44; Indels 17; Gaps 11;
QY 7 ttgttccagcaacaattatatatacacccctgaatcaacttaaggtgggtacaatt 66
|||||
Db 97 TTGGTTCAGCAACAATTTATATATATACACCCCTGAATCAACTTAAGGGTGTACAA-T 155
QY 67 gtcaatgtctatggtgtgtggaagttctttaagcccccatatctaaagcaagaaactgat 126
|||||
Db 156 GTCAATGTCTATGTTGTGAAGTTCTTTAAGCCCATATCTTAAGCAAGGAACGTAT 215
QY 127 tattgtcagttgtaactattgtgaccagacaataataaaactaaacttgcctgctctt 186
|||||
Db 216 TATTGCTCAGTTGTACTATTGTGGACGACAGCAAAATGTAAACTTAACCTTGCCTCTCTT 275
QY 187 agtggaaactatgaagcccttccaaataattataaaatggagatatattgttcgctttcac 246
|||||
Db 276 AGTGGAACTATGAAGCCCTTCCAATAATTATATAAATGGAGATATTGTTCGCTTTTAC 335
QY 247 aggtgaagattcaagtataaaaaagagagactcaggggtatcaccagctctggtttgca 306
|||||
Db 336 AGGCTGAAGATTCAAGTATATAAAGAGGAGACTCAGGGTATCACCAGCTCTGGCTTTGCA 395
QY 307 tctttgacgtttgaggaacttttggagccctcatatcaccttcgaccttcaagcaag-ta 365
|||||
Db 396 TCTTTGAGTTT-CAGGGAACCTTGGAGCCCTATCATACCTCGCACTTCACGCAAGCTA 453
QY 366 ttttaacttcaactactgaggaccacaaaatgttagaaccttcaagtgtttgggcatctac 425
|||||
Db 454 TTTTAAGTTCATCTACTGAGGACCAAAAATGGTAGAAGCCTTACGTG-TTGGGCACTAC 512
QY 426 tcatatgtcaccgtctggacattactaaaattgtgtgattgttcagccaatgcagattt 485
|||||

```
Query Match          6.7%; Score 87; DB 4; Length 10380;
Best Local Similarity 83.5%; Pred. No. 1.8e-17;
Matches 111; Conservative 0; Mismatches 20; Indels 2; Gaps 1;

QY 943 ttccaatgagatcctcctcagcagctgagttcagtcagtcgagtcgctc 1002
DB 6794 TTTGAGATAGATCTCGTCTGTCCACCCAGCTGGAGTGCAGTGCAGTCTCAGTTC 6853

QY 1003 attgcagctccaccctcctgagttcaagc--ttctcctgcctcagcctcccaagtgcg 1060
DB 6854 ACTGCAACCTCTGCTTCTGAGTTCAAGCGATTCTCTCTGCCTCAGCCTCCTGAGTAGCTG 6913

QY 1061 ggattacagctc 1073
DB 6914 GGATTACAGCGC 6926

RESULT 13
US-09-173-914-6
; Sequence 6, Application US/09173914
; Patent No. 6171857
; GENERAL INFORMATION:
; APPLICANT: Hendrickson, Eric
; TITLE OF INVENTION: A No. 6171857el Leucine Zipper, KARP-1 and
; TITLE OF INVENTION: Methods of Regulating DNA Dependent Protein Kinase Activity
; FILE REFERENCE: B0877/7017/HK
; CURRENT APPLICATION NUMBER: US/09/173,914
; CURRENT FILING DATE: 1998-10-16
; EARLIER APPLICATION NUMBER: 60/064,557
; EARLIER FILING DATE: 1997-10-17
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 14636
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-173-914-6

Query Match          6.7%; Score 87; DB 4; Length 14636;
Best Local Similarity 83.5%; Pred. No. 2.2e-17;
Matches 111; Conservative 0; Mismatches 20; Indels 2; Gaps 1;

QY 943 ttccaatgagatcctcctcagcagctgagttcagtcagtcgagtcgagtcgctc 1002
DB 8033 tttagacgagatcctcgtctgtcgtccagcagctgagtgaaagtgcgagatcgcgctc 8092

QY 1003 attgcagctccaccctcctcagttcaagc--ttctcctgcctcagcctcccaagtgcg 1060
DB 8093 actgcaactccgcctcccccgggttcaagctattctcctcagcctcccccagtagctg 8152

QY 1061 ggattacagctc 1073
DB 8153 ggattacagcgc 8165

RESULT 14
US-09-345-217-3/c
; Sequence 3, Application US/09345217
; Patent No. 6268142
; GENERAL INFORMATION:
; APPLICANT: DUFF, GORDON W.
; APPLICANT: COX, ANGELA
; APPLICANT: CAMP, NICOLA J.
; APPLICANT: DIGIOVINE, FRANCESCO S.
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR DISEASES ASSOCIATED
; TITLE OF INVENTION: WITH AN IL-1 INFLAMMATORY HAPLOTYPE
; FILE REFERENCE: MSA-010.02
; CURRENT APPLICATION NUMBER: US/09/345,217
; CURRENT FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: PCT/GB98/01481
; EARLIER FILING DATE: 1998-05-21
; EARLIER APPLICATION NUMBER: 9711040.7
```

```
; EARLIER FILING DATE: 1997-05-29
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 12565
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-345-217-3

Query Match          6.7%; Score 86.8; DB 4; Length 12565;
Best Local Similarity 85.2%; Pred. No. 2.3e-17;
Matches 109; Conservative 0; Mismatches 17; Indels 2; Gaps 1;

QY 946 ccaaatgagatcctcctcagcagctgagttcagtcagtcgagtcgagtcgctcatt 1005
DB 10822 CGAGACGAGTCTTGTCTGTGCGCTAGGCTGGAGTGCAGTGCAGTCTCGGCTCACT 10763

QY 1006 gcagcctccaccctcctcagttcaagc--ttctcctgcctcagcctcccaagtgcgga 1063
DB 10762 GCACCTCCACCTCCTGGTTCAAGCGATTCTCTCTGCCTCAGCCTCCTGAGTAGCTGGA 10703

QY 1064 ttacagcgc 1071
DB 10702 TTACAGGC 10695

RESULT 15
US-09-385-982-358
; Sequence 358, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 358
; LENGTH: 619
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(619)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-358

Query Match          6.7%; Score 86.6; DB 4; Length 619;
Best Local Similarity 84.0%; Pred. No. 4.3e-18;
Matches 110; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 943 ttccaatgagatcctcctcagcagctgagttcagtcagtcgagtcgagtcgctc 1002
DB 29 tttagagatgagatcctcgtctgtcgtccagcagctgagtcgagtcgagtcgctcgtc 88

QY 1003 attgcagctccaccctcctcagttcaagc--ttctcctgcctcagcctcccaagtgcg 1060
DB 89 actgcaactccgcctccctcgttcgaagcaattctcctcagcctccccaagtgcg 148

QY 1061 ggattacagcgc 1071
DB 149 ggattacagcgc 159
```

```

;
; FEATURE:
; NAME/KEY: intron
; LOCATION: 15116..15533
; OTHER INFORMATION: /number= 11
; FEATURE:
; NAME/KEY: exon
; LOCATION: 15534..15757
; OTHER INFORMATION: /number= 12
; FEATURE:
; NAME/KEY: intron
; LOCATION: 15758..16949
; OTHER INFORMATION: /number= 12
; FEATURE:
; NAME/KEY: exon
; LOCATION: 16950..17082
; OTHER INFORMATION: /number= 13
; FEATURE:
; NAME/KEY: intron
; LOCATION: 17083..17696
; OTHER INFORMATION: /number= 13
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17697..17764
; OTHER INFORMATION: /number= 14
; FEATURE:
; NAME/KEY: intron
; LOCATION: 17765..18534
; OTHER INFORMATION: /number= 14
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1776..1854, 2564..2621, 4076..4208,
; 6041..6252, 6802..6934, 7759..7856, 9444..9573,
; 10867..11081, 12481..12613, 13702..13799,
; 14977..15115, 15534..15757, 16950..17082,
; 17697..17741)
; OTHER INFORMATION: /product= "human serum albumin"
; OTHER INFORMATION: /citation= {[1]}
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18535..18697
; OTHER INFORMATION: /number= 15
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 17742..18697
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1737..1775
; LOCATION INFORMATION:
; AUTHORS: Minghetti, P P
; AUTHORS: Ruffner, D E
; AUTHORS: Kuang, W-J
; AUTHORS: Dennison, O E
; AUTHORS: Hawkins, J W
; AUTHORS: Beattie, W G
; AUTHORS: Dugaiczkyk, A
; TITLE: MOLECULAR STRUCTURE OF THE HUMAN ALBUMIN
; TITLE: GENE IS REVEALED BY NUCLEOTIDE SEQUENCE WITHIN
; TITLE: q11-22 OF CHROMOSOME 4
; JOURNAL: J. Biol. Chem.
; VOLUME: 261
; PAGES: 6747-6757
; DATE: 1986
; RELEVANT RESIDUES IN SEQ ID NO: 36: FROM 1 TO 19011
;
US-08-310-356-36
```

Query Match 6.7%; Score 87.2; DB 1; Length 19011;
Best Local Similarity 84.6%; Pred. No. 2.2e-17;
Matches 110; Conservative 0; Mismatches 18; Indels 2; Gaps 1;

```

Oy 943 ttccaaatgagctcgttcctccaggtgagttcagtgccagggtcagctcgcctc 1002
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15309 TTTTGACGCGAGTCTCGCTTTGTTGTCAGGCTGGAGTCAGTGGTCCATCTCGCTC 15368
```

```

Oy 1003 attgcagctccacctcctgagttcaagc--ttctctgcctccagctcccaagtagctg 1060
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15369 ACTGCAACCTCCGCTCCCAAGTTCAAGCATTCCTCGCTCAGCTCCCAAGTAGCTG 15428

Oy 1061 ggattacagg 1070
   ||| ||| ||| |||
Db 15429 GGACTACAGG 15438
```

```

RESULT 10
PCT-US92-06300-1
; Sequence 1, Application PC/TUS9206300
; GENERAL INFORMATION:
; APPLICANT: Hurwitz, David R
; APPLICANT: Nathan, Margret
; APPLICANT: Shani, Moshe
; TITLE OF INVENTION: Transgenic Protein Production
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer, Inc.
; STREET: 500 Virginia Ave., Bldg. 3A
; CITY: Ft. Washington
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19034
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06300
; FILING DATE: 19920730
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Goodman, Rosanne
; REGISTRATION NUMBER: 52,534
; REFERENCE/DOCKET NUMBER: A0856-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 962-4130
; TELEFAX: (215) 962-4107
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19557 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PUBLICATION INFORMATION:
; AUTHORS: Minghetti, P P
; AUTHORS: Ruffner, D E
; AUTHORS: Kuang, W.-J.
; AUTHORS: Dennison, O E
; AUTHORS: Hawkins, J W
; AUTHORS: Beattie, W G
; AUTHORS: Dugaiczkyk, A
; TITLE: MOLECULAR STRUCTURE OF THE HUMAN ALBUMIN GENE
; TITLE: IS REVEALED BY NUCLEOTIDE SEQUENCE WITHIN Q11-22
; TITLE: OF CHROMOSOME 4
; JOURNAL: J. Biol. Chem.
; VOLUME: 261
; PAGES: 6747-6757
; DATE: 1986
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 19002
;
PCT-US92-06300-1
```

Query Match 6.7%; Score 87.2; DB 5; Length 19557;
Best Local Similarity 84.6%; Pred. No. 2.2e-17;
Matches 110; Conservative 0; Mismatches 18; Indels 2; Gaps 1;

Qy	943	tttccaaatggagctctcgcttcagctctccagctctgagatctcagtgacagctctcggtc	1002
Db	8727	TTTTGACGAGGATCTCGTTCTGTACCCGCGCTGGAGTGCAGTGCACATCTCGGCTC	8668
Qy	1003	attcgagcctccacctctctgagttcaagc--ttctctgctctcagctctcccaagtagctg	1060
Db	8667	ACTGCAACTCCGCCCTCCGGGTTCAAGCTATTCTCTGCCTCAGCCCTCCCAAGTAGCTG	8608
Qy	1061	ggattacaggctct	1074
Db	8607	GGACTACAGGCGCT	8594

	Query Match	Best Local Similarity	84.3%	6.9%	Score 89.6;	DB 2:	Length 246240;
	Matches 113;	Conservative	17;	Mismatches	19;	Indels	2: Gaps
y	943	tttccaaatggagctctcgctcttcgctccacgagctgagttcagtgacagcgtctctcggtc	1002				
b	8727	TTTTTGACGGAGCTCTCGTTCGTCTGTCACCCAGGCTGGAGTGCAGTGGCACAAATCTCGGCTC	8668				
y	1003	atcgagcgtccaccctctcgtgagttcaagc--ttctcctgcctcagcctcccaagtagctg	1060				
b	8667	actcgagctctccgctctccgggttcaagctatttctctgcttcagcctcccaagttagctg	8608				

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-07754A-2

Query Match 3.7%; Score 122; DB 5; Length 1144;
Best Local Similarity 18.2%; Pred. No. 0.0039;
Matches 135; Conservative 106; Mismatches 200; Indels 300; Gaps 36;

Qy 71 IYKNGDIVRFHRL----KIOVYKKEQ--ITSSGFASLTFFCTLCAPITPRTSSKYN 124
Db 335 LIEKNDIIEVTALPDHESIQLFQKHAQFKEVNEFKLSLEVNYAKGLP----- 387

Qy 125 FTTEHKKVREALRWAS-----THM---SPSWTLKL-----CDVQPMQ--- 160
Db 388 -----ALKVWGSLHLNRLTEWKSIAEHMKNNSYSGIIDKLSIDGLEPKQEM 437

Qy 161 YFDLTCLLGAEDVGSFLKKVWD-----GTRTPPFSWRVLQDLVLEGLSHIR 212
Db 438 FLDIACFLRGEK----DYILOLESCHIGAEGYGLR-----ILIDKSLVFISE 481

Qy 213 LQNLTDILVYDNHVRVARSK-VGSFLRIYSLHTKLQSMNSNOTMLSLEFHLHGTSY 271
Db 482 YNOVOMHDLQDMGKIYVNFQKDGERSRLW-LAKEVEEVMNSNTGTMAE--AIWVSSY 538

Qy 272 GRGIRVLPESNDVDQLKDLSEANL-TANOHSV-----ICQEPDDSPSSSGS 320
Db 539 SSTLRF---SNOAVKMKR-LRVFNMGSRSTHYAIDLPLNNLCFCVCTNYPWESFPSTFE 594

Qy 321 -----VSELYEVER 328
Db 595 LKMLVHLQLRHNSLRHLWTEKHLPLSLRRIDLSWSKRLTRPTDFTGMPNLEYVNYOCSN 654

Qy 329 -----COOLSATILTDHOYLERTPLCAILK-----OKAPOY-RIRA 364
Db 655 LEEVHSLGCCSKVIGLYLNDCKSLKRP-CVNVESLEYLGLRSCDSLEKLPYGRMKP 713

Qy 365 KLRYSK-----PRLFOSVKLHCPK-----CHL 387
Db 714 EIQIHMOGSGIRELPSSIFQ-YKTHVTKLLNNKNLVALPSSICRLKLSVLSVSCSK 772

Qy 388 LOEVPHE-GDLDIIFODGATKTPVYKQNTSLYDSKLTWTKNOKGRVAVHFKVNGILP 446
Db 773 LESLPEEIGDLNLRVFDASDTLILRPPSSIIRLNLK-IILMFRGKDGVHF----- 823

Qy 447 LSNBCLLLIEGGTLEICKLSKNFNSVIPVRSGHDELELDLSAPFLIOGTHHYGCKQC 506
Db 824 -----EFPVVAEGLHSLNLSYCNLIDGSLP-----EEI 854

Qy 507 SSLRSIQNLNLSVDKTSWIPSSVAEALGIPLQVYVMTFLDDGTGVLEAY-LMDSKDF 565
Db 855 GLSLKLKDLNLRNFEHLPSIAQ-----LGALQSLDKDCORL 894

Qy 566 FOIPASEVLMDLQKSDVMDMFCPPGIK---IDAYPWLECFIKSYNYI----- 613
Db 895 TQLEP-----LPPELNLHVDCHEMALK-FI-HYLVTKRKLHRVK 932

Qy 614 -NGTDNQCIOYIFDTTVAEDV 633
Db 933 LDDAHNDMTYNLFAYTMFQNI 953

RESULT 8
PCT-US95-07754A-4
; Sequence 4, Application PC/TUS9507754A
; GENERAL INFORMATION:
; APPLICANT: Baker, Barbara J
; APPLICANT: Whitham, Steven A
; TITLE OF INVENTION: Plant Virus Resistance Gene and Methods
; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Margaret A. Connor, USDA-ARS-
; STREET: 800 Buchanan Street
; CITY: Albany
; STATE: CA
; COUNTRY: USA
; ZIP: 94710
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07754A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Connor, Margaret A
; REGISTRATION NUMBER: 30043
; REFERENCE/DOCKET NUMBER: 0094.94
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 559-6067
; TELEFAX: (510) 559-5777
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-07754A-4

Query Match 3.7%; Score 122; DB 5; Length 1144;
Best Local Similarity 18.2%; Pred. No. 0.0039;
Matches 135; Conservative 106; Mismatches 200; Indels 300; Gaps 36;

Qy 71 IYKNGDIVRFHRL----KIOVYKKEQ--ITSSGFASLTFFCTLCAPITPRTSSKYN 124
Db 335 LIEKNDIIEVTALPDHESIQLFQKHAQFKEVNEFKLSLEVNYAKGLP----- 387

Qy 125 FTTEHKKVREALRWAS-----THM---SPSWTLKL-----CDVQPMQ--- 160
Db 388 -----ALKVWGSLHLNRLTEWKSIAEHMKNNSYSGIIDKLSIDGLEPKQEM 437

Qy 161 YFDLTCLLGAEDVGSFLKKVWD-----GTRTPPFSWRVLQDLVLEGLSHIR 212
Db 438 FLDIACFLRGEK----DYILOLESCHIGAEGYGLR-----ILIDKSLVFISE 481

Qy 213 LQNLTDILVYDNHVRVARSK-VGSFLRIYSLHTKLQSMNSNOTMLSLEFHLHGTSY 271
Db 482 YNOVOMHDLQDMGKIYVNFQKDGERSRLW-LAKEVEEVMNSNTGTMAE--AIWVSSY 538

Qy 272 GRGIRVLPESNDVDQLKDLSEANL-TANOHSV-----ICQEPDDSPSSSGS 320
Db 539 SSTLRF---SNOAVKMKR-LRVFNMGSRSTHYAIDLPLNNLCFCVCTNYPWESFPSTFE 594

Qy 321 -----VSELYEVER 328
Db 595 LKMLVHLQLRHNSLRHLWTEKHLPLSLRRIDLSWSKRLTRPTDFTGMPNLEYVNYOCSN 654

Qy 329 -----COOLSATILTDHOYLERTPLCAILK-----OKAPOY-RIRA 364
Db 655 LEEVHSLGCCSKVIGLYLNDCKSLKRP-CVNVESLEYLGLRSCDSLEKLPYGRMKP 713

Qy 365 KLRYSK-----PRLFOSVKLHCPK-----CHL 387
Db 714 EIQIHMOGSGIRELPSSIFQ-YKTHVTKLLNNKNLVALPSSICRLKLSVLSVSCSK 772

Qy 388 LOEVPHE-GDLDIIFODGATKTPVYKQNTSLYDSKLTWTKNOKGRVAVHFKVNGILP 446
Db 773 LESLPEEIGDLNLRVFDASDTLILRPPSSIIRLNLK-IILMFRGKDGVHF----- 823

Qy 447 LSNBCLLLIEGGTLEICKLSKNFNSVIPVRSGHDELELDLSAPFLIOGTHHYGCKQC 506

LENGTH: 1143 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-04589-108

Query Match 3.7%; Score 122; DB 5; Length 1143;
Best Local Similarity 18.2%; Pred. No. 0.0039;
Matches 135; Conservative 106; Mismatches 200; Indels 300; Gaps 36;

```
OY 71 IYKNGDIVRFHRL-----KIOYKKEG--ITSSGFASLTGEGTLGAPIIPRTSSKYFN 124
Db 335 LIEKNDIIEVTALPDHESIQLFQHAFGKEVPNENFEKLSLEVNVYAKGLPL----- 387
OY 125 FTTEHKKVVEALRWAS-----THM-----SPSWTLKL-----CDVQPMQ--- 160
Db 388 -----ALKVWGSLHLNRLTEWKAIEHMKNNYSYGIIDKLSYDGLPKQOEM 437
OY 161 YFDLTCLGKAEVDGASFLKLVMD-----GTRIPPSNRVLQDLVLEGDLSHHR 212
Db 438 FLDIACFLGEEK-----DYILOLESCHIGAEGYGLR-----ILDKSLVFISE 481
OY 213 LQNLITDILVDNHRVHVARSLK-VGSFLRIYSLHTKLSQMSNENOTMLSEFHLHGTSY 271
Db 482 YNQVQMHDLQDMGKYIVNFQKDGERSRLW-LAKEVEEVMNSNNTGTNAME--AIWVSSY 538
OY 272 GRGTRVLPESNSVDOLKDDLESANL-TANOHSOV-----ICQSEPDSPSSGS 320
Db 539 SSTLRF---SNOAVKNMKR-LRVFNMGSRSTHYAIDYLPNNLRFCVCTNYPWESFPSTFE 594
OY 321 -----PRRLFQSVKLHCPC-----VSLYEVEVER 328
Db 595 LKMLVHLQLRHNSLRHLWTETKHLPSLRIDLSWSKRLTRPDTGMPNLEYVNLQCSN 654
OY 329 -----COOLSATILTDHOYLERTPLCAILK-----OKAPOQY-RIRA 364
Db 655 LEEVHSLGCKSKVIGLYLNDCKSLKRF-CVNVESLEYLGLRSCDSLEKLPYIGRMKP 713
OY 365 KLRYSK-----PRRLFQSVKLHCPC-----CHL 387
Db 714 EIOIHMOGSGIRELPSIFQ-YKTHVTKLLWNMKNLVALPSSICRLKSLVLSVSGSK 772
OY 388 LOEVPHE-GDLTIIFODGATKTPVVKLQNTSLYDSKIWTTRKQKRVAVHVKNGILP 446
Db 773 LESLPEEIGDLNLRVFDASDTLILRPPSSIIRLNKL-IILMFRGKDGVF----- 823
OY 447 LSNBCLLLIEGGTLSEICKLSKNFNSVIPVRSGHEDLELDSAPFLIOGTIHHYCKQC 506
Db 824 -----EFPVVAEGLHSLEYLNLSYCNLIDGGLP-----EEI 854
OY 507 SSLRSIONLSYDKTSWIPSSVAEALGIVPLQYVFMFTLDDGTGVLEAY-LMDSKRF 565
Db 855 GSLSLKLKDLNRSNFEHLPSIAQ-----LQALQSLDKDCQRL 894
OY 566 FOIPASEVLMDLQKSDVHMDMFCPPGK---IDAYPMLECFIKSYNVT----- 613
Db 895 TOLPE-----LPPELNLHVDCHMAUK-FI-HYLVTRKKKLHRVK 932
OY 614 -NGTDNQCICQIFDTTVAEDV 633
Db 933 LDDAHNDTMYNLFAYTMFQNI 953
```

RESULT 5

US-08-261-663A-2

Sequence 2, Application US/08261663A

Patent No. 5571706

GENERAL INFORMATION:

APPLICANT: Baker, Barbara J

APPLCANT: Whitham, Steven A

TITLE OF INVENTION: Plant Virus Resistance Gene and Methods

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:
ADDRESSEE: Margaret A. Connor, USDA-ARS
STREET: 800 Buchanan Street
CITY: Albany
STATE: CA
COUNTRY: USA
ZIP: 94710
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261.663A
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Connor, Margaret A
REGISTRATION NUMBER: 30043
REFERENCE/DOCKET NUMBER: 0094.94
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 559-6067
TELEFAX: (510) 559-5777
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1144 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-261-663A-2

Query Match 3.7%; Score 122; DB 1; Length 1144;

Best Local Similarity 18.2%; Pred. No. 0.0039;

Matches 135; Conservative 106; Mismatches 200; Indels 300; Gaps 36;

```
OY 71 IYKNGDIVRFHRL-----KIOYKKEG--ITSSGFASLTGEGTLGAPIIPRTSSKYFN 124
Db 335 LIEKNDIIEVTALPDHESIQLFQHAFGKEVPNENFEKLSLEVNVYAKGLPL----- 387
OY 125 FTTEHKKVVEALRWAS-----THM-----SPSWTLKL-----CDVQPMQ--- 160
Db 388 -----ALKVWGSLHLNRLTEWKAIEHMKNNYSYGIIDKLSYDGLPKQOEM 437
OY 161 YFDLTCLGKAEVDGASFLKLVMD-----GTRIPPSNRVLQDLVLEGDLSHHR 212
Db 438 FLDIACFLGEEK-----DYILOLESCHIGAEGYGLR-----ILDKSLVFISE 481
OY 213 LQNLITDILVDNHRVHVARSLK-VGSFLRIYSLHTKLSQMSNENOTMLSEFHLHGTSY 271
Db 482 YNQVQMHDLQDMGKYIVNFQKDGERSRLW-LAKEVEEVMNSNNTGTNAME--AIWVSSY 538
OY 272 GRGTRVLPESNSVDOLKDDLESANL-TANOHSOV-----ICQSEPDSPSSGS 320
Db 539 SSTLRF---SNOAVKNMKR-LRVFNMGSRSTHYAIDYLPNNLRFCVCTNYPWESFPSTFE 594
OY 321 -----VSLYEVEVER 328
Db 595 LKMLVHLQLRHNSLRHLWTETKHLPSLRIDLSWSKRLTRPDTGMPNLEYVNLQCSN 654
OY 329 -----COOLSATILTDHOYLERTPLCAILK-----OKAPOQY-RIRA 364
Db 655 LEEVHSLGCKSKVIGLYLNDCKSLKRF-CVNVESLEYLGLRSCDSLEKLPYIGRMKP 713
OY 365 KLRYSK-----PRRLFQSVKLHCPC-----CHL 387
Db 714 EIOIHMOGSGIRELPSIFQ-YKTHVTKLLWNMKNLVALPSSICRLKSLVLSVSGSK 772
OY 388 LOEVPHE-GDLTIIFODGATKTPVVKLQNTSLYDSKIWTTRKQKRVAVHVKNGILP 446
Db 773 LESLPEEIGDLNLRVFDASDTLILRPPSSIIRLNKL-IILMFRGKDGVF----- 823
OY 447 LSNBCLLLIEGGTLSEICKLSKNFNSVIPVRSGHEDLELDSAPFLIOGTIHHYCKQC 506
```

Db 335 LIEKNDIIEVTALPDHESIQLFKOHAFGEVNPNEFKLSLEVYVYAKGLPL----- 387
QY 125 FTEDHKMVEALRVAS-----THM-----SPSWTLKL-----CDVOPMO--- 160
Db 388 -----ALKVWGSLLHNLRLTEKSAIEHKMKNYSIGIDKLSISYDGLPEKQOEM 437
QY 161 YFDLTCLLGAEDVGFASFLKVD-----GTRTPFSPSRVLIODLVLEGLSHIHR 212
Db 438 FLDIACFLRGEK-----DYILOLESCHIGAEGYLR-----ILDKSLVFISE 481
QY 213 LONITIDILVYDNHVVHVARSLK-VGSFLRIYSLHTKLOSMNSNOTMLSLEFLHGGTSY 271
Db 482 YNOVOMHDLQDMGKIYVNFQKPCERSRLW-LAKEVEEYMSNNTGTMAEIM--VSSY 538
QY 272 GRGIRVLPESNDVDOLKDLKLESANL-TANOHSV-----ICOSEPDDSPSSGS 320
Db 539 SSTLRP---SNOAVKNMKR-LRVFNMGSRSTHYAIDYLPNNLRFCVCTNYPWESFPSTFE 594
QY 321 -----PRRLFOSVKLHCPK-----VSLYEYER 328
Db 595 LKMLVHLQRLHNSLRHLWTETKHLPSLRRLIDLSNKRTRTPDFTGMPNLEVYVNLQCSN 654
QY 329 -----COOLSATILTDHOYLERTPLCALK-----OKAPOOY-RIRA 364
Db 655 LEEVHSLGCCSKVIGLYLNDCKSLKRF-CVNVESLEYLGLRSCDSLEKLEPEIYGRMKP 713
QY 365 KLSRYK-----PRRLFOSVKLHCPK-----CHL 387
Db 714 EQIHMGGSGIRELPSISFO-YKHTVKLLNMKNLVALPSSICRLKSLVSLVSGCSK 772
QY 388 LOEVPEH-GDLDIIFODGATKTPVVKLQNTSLYDSKIWTTRKNOKRKVAVHFKVNGILP 446
Db 773 LESLPEEGDLDNLRVFDASDTLIRPPSSIIRLNKL-IILMFRGFKDGVHF----- 823
QY 447 LSNCELLIEGTLSEICKLSNKNFNSVIPVRSCHDELLELDLSAPFLIOGTHHYGKQC 506
Db 824 -----EPPVAEGLSHLEYLNLSCNLDGGLP-----EEI 854
QY 507 SSLRSIONLSLVKTSWIPSSVAEALGIVPQYVFMTFTLDDGTGVLEAY-LMDSKX 565
Db 855 GSLSLKLRLSRNFEHLPSIAQ-----LGAOSLDLKCQRL 894
QY 566 FQIPASEVLDLDDLOKSDVMDMFCPPGK-----IDAYPWLECFIKSYNYT----- 613
Db 895 TOLPE-----LPPELNLHVDCHMALK-FI-HYLVTKRKKLHRVK 932
QY 614 -NGTDNOICQIIFDITVAEDV 633
Db 933 LDDAHNDTHNLFAYTMFQNI 953

RESULT 2
US-08-310-912A-108
: Sequence 108, Application US/08310912A
: Patent No. 5981730
: GENERAL INFORMATION:
: APPLICANT: Ausubel, Frederick M.
: APPLICANT: Staskawicz, Brian J.
: APPLICANT: Brent, Andrew F.
: APPLICANT: Dahlbeck, Douglas
: APPLICANT: Katagiri, Fumitaki
: APPLICANT: Kunkel, Barbara N.
: APPLICANT: Mindrinos, Michael N.
: APPLICANT: Yu, Guo-Liang
: TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND DETECTION
: NUMBER OF INVENTION: METHODS
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA

: COUNTRY: USA
: ZIP: 02110-2904
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: FILING DATE: September 22, 1994
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/227,360
: FILING DATE: April 13, 1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Lech, Karen F.
: REGISTRATION NUMBER: 35,238
: REFERENCE/DOCKET NUMBER: 00786/254001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 542-5070
: TELEFAX: (617) 542-8906
: TELEX: 100254
: INFORMATION FOR SEQ ID NO: 108:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1143 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-310-912A-108

Query Match 3.7%; Score 122; DB 2; Length 1143;
Best Local Similarity 18.2%; Pred. No. 0.0039;
Matches 135; Conservative 106; Mismatches 200; Indels 300; Gaps 36;

QY 71 IYKNGDIVRFLR-----KIQVYKKEQOG--ITSSGFASLTFTGLGAPIIPTSKSYFN 124
Db 335 LIEKNDIIEVTALPDHESIQLFKOHAFGEVNPNEFKLSLEVYVYAKGLPL----- 387
QY 125 FTEDHKMVEALRVAS-----THM-----SPSWTLKL-----CDVOPMO--- 160
Db 388 -----ALKVWGSLLHNLRLTEKSAIEHKMKNYSIGIDKLSISYDGLPEKQOEM 437
QY 161 YFDLTCLLGAEDVGFASFLKVD-----GTRTPFSPSRVLIODLVLEGLSHIHR 212
Db 438 FLDIACFLRGEK-----DYILOLESCHIGAEGYLR-----ILDKSLVFISE 481
QY 213 LONITIDILVYDNHVVHVARSLK-VGSFLRIYSLHTKLOSMNSNOTMLSLEFLHGGTSY 271
Db 482 YNOVOMHDLQDMGKIYVNFQKPCERSRLW-LAKEVEEYMSNNTGTMAEIM--VSSY 538
QY 272 GRGIRVLPESNDVDOLKDLKLESANL-TANOHSV-----ICOSEPDDSPSSGS 320
Db 539 SSTLRP---SNOAVKNMKR-LRVFNMGSRSTHYAIDYLPNNLRFCVCTNYPWESFPSTFE 594
QY 321 -----PRRLFOSVKLHCPK-----VSLYEYER 328
Db 595 LKMLVHLQRLHNSLRHLWTETKHLPSLRRLIDLSNKRTRTPDFTGMPNLEVYVNLQCSN 654
QY 329 -----COOLSATILTDHOYLERTPLCALK-----OKAPOOY-RIRA 364
Db 655 LEEVHSLGCCSKVIGLYLNDCKSLKRF-CVNVESLEYLGLRSCDSLEKLEPEIYGRMKP 713
QY 365 KLSRYK-----PRRLFOSVKLHCPK-----CHL 387
Db 714 EQIHMGGSGIRELPSISFO-YKHTVKLLNMKNLVALPSSICRLKSLVSLVSGCSK 772
QY 388 LOEVPEH-GDLDIIFODGATKTPVVKLQNTSLYDSKIWTTRKNOKRKVAVHFKVNGILP 446
Db 773 LESLPEEGDLDNLRVFDASDTLIRPPSSIIRLNKL-IILMFRGFKDGVHF----- 823
QY 447 LSNCELLIEGTLSEICKLSNKNFNSVIPVRSCHDELLELDLSAPFLIOGTHHYGKQC 506

RC TISSUE-Petal brain;
RX MEDLINE-99056138; PubMed-9806842;
RA Fjiwara T., Saito A., Suzuki M., Shinomiya H., Suzuki T.,
RA Takahashi E., Tanigami A., Ichiiyama A., Chung C.H., Nakamura Y.,
RA Tanaka K.;
RT "Identification and chromosomal assignment of USP1, a novel gene
RL encoding a human ubiquitin-specific protease.";
RL Genomics 54:155-158(1998).
RN [2]
RN SEQUENCE FROM N.A.
RC TISSUE-Testis;
RX MEDLINE-21154917; PubMed-11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansorge W., Boecker M., Bloecker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Hewes H.W., Ottenwelder B., Obermaier B., Rampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
CC -1- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O =
CC UBIQUITIN + A THIOL.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS
CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB014458; BAA34703.1; -
DR EMBL: AL117575; CAB55999.1; -
DR EMBL: AL117503; CAB55967.1; -
DR MEROPS: C19.019; -
DR MIM: 603478; -
DR InterPro: IPR001394; UCH-2.
DR Pfam: PF00442; UCH-1; 1.
DR Pfam: PF00443; UCH-2; 1.
DR PROSITE: PS00972; UCH_2_1; 1.
DR PROSITE: PS00973; UCH_2_2; 1.
DR PROSITE: PS0235; UCH_2_3; 1.
DR K/F Ubiquitin conjugation; Hydrolase; Multigene family.
DR ACT_SITE 90 90 BY SIMILARITY.
DR ACT_SITE 584 584 BY SIMILARITY.
DR ACT_SITE 593 593 BY SIMILARITY.
DR ACT_SITE 593 593 BY SIMILARITY.
DR SEQUENCE 785 AA; 88207 MW; 50AA2817A60810AF CRC64;

Query Match 3.1% Score 104; DB 1; Length 785;
Best Local Similarity 19.8% Pred. No. 5.9;
Matches 96; Conservative 64; Mismatches 213; Indels 112; Gaps 18;
QY 154 CDVQPMQVFDLTC---QLLGRAEVGASFLKLVGDGTRTPFSPNRVLQDL-----VLEG 205
DB 135 KREDSLASYELICSLSQSLIIISVEQLQASFLNPKYKIDELATQPRRLNTLRELNPMYEG 194
QY 206 DLSHTRHLQNLITDIL--VYDNHVRVARSCLKVGSFLRIYSLHTKLOSNSENQ:MLSLEF 263
DB 195 YLQHDQAQ-----EVLCILGNIQETCOLLAKKEVKNVAELTPKEIPHPKHEM----- 243
QY 264 HLHGSTYGRGIRVLPSNSVDVQLKDDLESANLTANOHSVICQSEPDOSFSGSVSL 323
DB 244 -----NGINSIEMDSMRHSEDFKEKLPKGNKFKSDTEF 277
QY 324 YEVEPCQOLSATILTDHOYLERTPLCALLKOKAQOQYIRAKLSYKPRRLFQSVKHLCP 383
DB 278 GNMKKKVKLS-----KEHOSLEENORQTRSKRKATSD-----TLESPPKTIIP 319
QY 384 KCHLLQEVPH-----EGDLIDIIFOGATKTPVVKLQNTSLYDSKIWTITKNQGRVAVHFV 439

DB 320 KYISENPRSPQSKSRKVINLWLSATKQPSILLSKFCSL--GKITNQGVKGQS----- 371
QY 440 KNGILP-----LSNECLLLIEGGTSLSEICKLSNKFNSVIVRSGHEDL--ELL 486
DB 372 KENCDEPDEDLGKCESDNTTNGCGLESPGNTVTPV-----NVNEVKPKNGEQIGFELY 426
QY 487 DLSAPELIQO-TIHHYCKKCCSSL-----RSTONLSLVDKTSTWIPSSVAEALGIVPLQYV 541
DB 427 EK-----LFQGLVLRTRCLECESLTERDEFQDISVPVQDEL--SKVESSEISPEPKT 480
QY 542 FVMTFTLDDGTGVLEAYLMDSDKFF-----QIPASEVLMDDDLQKSVDMIMDMFCPPGI 595
DB 481 EMKTLRWAIISQFASVERIVGDKYVCENCCHHYTEAERSLLFLDKMPEVITTHLKCFASAGL 540
QY 59% KIDAY 500
DB 541 EFCY 545

RESULT 14
UVRA_CHLMU
ID UVRA_CHLMU STANDARD; PRT; 1787 AA.
AC Q9PK60;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE EXINUCLEASE ABC SUBUNIT A.
GN UVRA OR TC0610.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MoPn / Nigg.
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Elsen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae Ar39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
CC CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
CC PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). UVRA IS AN ATPASE
CC AND A DNA-BINDING PROTEIN THAT PREFERENTIALLY BINDS SINGLE-
CC STRANDED OR UV-IRRADIATED DOUBLE-STRANDED DNA (BY SIMILARITY).
CC -1- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRA, UVRB AND UVRC.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). CONTAINS FOUR ABC DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE002329; AAF39441.1; -
DR TIGR: TC0610; -
DR InterPro: IPR003439; ABC_transportr.
DR Pfam: PF00005; ABC_tran; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW SOS response; Excision nuclease; DNA repair; ATP-binding; Repeat;
KW DNA-binding; Zinc-finger; Complete proteome.
FT NP_BIND 32 39 ATP (POTENTIAL).
FT NP_BIND 625 632 ATP (POTENTIAL).
FT NP_BIND 964 971 ATP (POTENTIAL).
FT NP_BIND 1509 1516 ATP (POTENTIAL).
FT ZN_FING 719 742 C4-TYPE.

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; J03358; AAA61190.1; -;
DR PIR; A31943; TVHUF6.
DR HSSP; P11362; 1FGI.
DR MLM; 176942; -;
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001050; FCH.
DR InterPro; IPR000980; SH2.
DR InterPro; IPR001245; Tyr_kin.
DR Pfam; PF00611; FCH; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00017; SH2; 1.
DR PRINTS; PR00109; TYRKINASE.
DR SMART; SM00055; FCH; 1.
DR SMART; SM00252; SH2; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00001; SH2; 1.
KW Transferase; Tyrosine-protein kinase; Proto-oncogene; ATP-binding;
KW Phosphorylation; Nuclear protein; SH2 domain; Polymorphism.
FT DOMAIN 460 550
FT NP_BIND 563 816 PROTEIN KINASE.
FT BINDING 569 577 ATP (BY SIMILARITY).
FT ACT_SITE 591 591 ATP (BY SIMILARITY).
FT MOD_RES 684 684 BY SIMILARITY.
FT VARIANT 714 714 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT V -> L.
FT /FTID=VAR_006282.
SQ SEQUENCE 822 AA: 94623 MW: 0491CD6392DB415 CRC64;

Query Match 3.28; Score 107; DB 1; Length 822;
Best Local Similarity 20.08; Pred. No. 3.8;
Matches 127; Conservative 82; Mismatches 213; Indels 214; Gaps 29;

Oy 37 LSKGTDYCVYTIYDQTNVLTCLLFGSGNEALPIYK-----NGDIVRFHRLKIQVY 89
Db 49 LCNQVDESTVOMYVSVNSKSWLLMIQOTEQLSRIMKTHAEDLNSGFL---HRLTMMI- 104
Oy 90 KKEQTITSSGFSALTEGLGAPIIPTSS-----KYFNFTEDHKVVEAL 136
Db 105 -KDKQVKKK---YIGVHQIQAEMIKVTKTELEKLCYSRQLIKEMNSAKERYK--EAL 158
Oy 137 RWAST-----HMSPSWTLKL--CDVOPMQYFDLTCLGKAEVDGASFL 180
Db 159 AKGRETEKAKERYDKATMKLHMLHNVYALKGAOLHONQYDITLPLLDLSLOKMOEM 218
Oy 181 LKWDGTRTPFPFMRVLQDLVLEGLSHIRLQNLITDIL-----YVDNHVHVARSLKVG 236
Db 219 IKALKGIFDEYSQ-----ITSLVTE-EIVNVHKEIOMSVQEQIDPSTEXNFDVHRITTA-- 271
Oy 237 SFLRIYSLHTKLQ-----SNSENQTMLSLEFHLHGGTSYGRGI-----RV 277
Db 272 -----AKEQEIEFTDSLLEENENIQAENEIMMNNLTAEISLQVLMKTLAEELMOTQOM 322
Oy 278 LPESNSDVLQKDKLESANLTANQSHSDVICOSEPDGFPSPSGSVSLVEVERCOOLSATIL 337
Db 322 LNKREAVLELEKRIEESSECEKKSQIV-----LLL 354
Oy 338 TDHGYLRTPLCALIKAKAPQOYRIKRLSKYKPRRLFSQVKLHCPKCHLLQVPEHGDL 397
Db 355 SOKQALEE-----LKQSVQOOLRTEAKFSQAK-----ELLEQKVOE--- 390
Oy 398 DIIFDQATKTPVVKLQNTSLYDSKIWTTKMQRKQVAVHFVKN--GIL--PLSNECLL 453

Db 391 ----NDGKEPPPVVYEE-----DARSVTSMERKLSKFESIRHSTAGIIRSPKS----- 437
Oy 454 LIEGGTLSEICKLSNK-----FNSVIIPVRSGHEDLELLDLSAPFLI----- 494
Db 438 AVGSALSMDISISEKPLAEODWYHGAIP---RIEAQELLKKQGDPLVRESHGKPGEYVL 494
Oy 495 -----OGT-----IHGYCKOCSSLSR-IQNLSILVDT 522
Db 495 SVYSDGQRHFIIQYVDNMVRFEGTGFNSIPQLIDHHYTKQVITKKSGVLLNPIPKDK 554
Oy 523 SWIPSS-----VAEALGIVLPQYVYVMTFTLDGTV 554
Db 555 KWILSHEDVILGELLG--KGNFGEVYKGTLDKDKTSV 588
RESULT 11
ID SP2E_BACME STANDARD; PRT; 585 AA.
AC P49600;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE STAGE II SPOULATION PROTEIN E (EC 3.1.3.16) (FRAGMENT).
GN SPOIIE.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1404;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=VT1660;
RX MEDLINE=96249699; PubMed=8830262;
RA Barak I., Behari J., Olmedo G., Guzman P., Brown D.P., Castro E.,
Walker D., Westpheling J., Youngman P.;
RT Structure and function of the Bacillus SpoIIIE protein and its
RT localization to sites of sporulation septum assembly.";
RL Mol. Microbiol. 19:1047-1060(1996).
CC -!- FUNCTION: NORMALLY NEEDED FOR PRO-SIGMA E PROCESSING DURING
CC SPOULATION BUT CAN BE BYPASSED IN VEGETATIVE CELLS. ACTIVATES
CC SPOIIIA BY DEPHOSPHORYLATION (BY SIMILARITY).
CC CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN +
CC ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. POLAR SEPTUM (BY
CC SIMILARITY).
CC -!- SIMILARITY: CONTAINS A PP2C-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; U26836; AAB58072.1; -;
DR InterPro; IPR003589; PP2C_catalytic.
DR InterPro; IPR003588; PP2C_sig.
DR SMART; SM00332; PP2C; 1.
DR SMART; SM00331; PP2C_SIG; 1.
KW Sporulation; Hydrolase; Transmembrane.
FT NON_TER 1
FT TRANSHEM 40 57 POTENTIAL.
FT TRANSHEM 70 86 POTENTIAL.
FT DOMAIN 366 585 PP2C-LIKE.
SQ SEQUENCE 585 AA: 65690 MW: 90E9ACF1D3E21D01 CRC64;

Query Match 3.28; Score 105; DB 1; Length 585;

Best Local Similarity 19.48; Pred. No. 3.2;
Matches 120; Conservative 89; Mismatches 202; Indels 208; Gaps 29;
Oy 46 VVT--IVDOTNW-----KLTCLLFGSGNEALPIYKNGDIVRFH-----RLKIQVYKKT 93


```

Db 368 TOF--YITRVPDVKEVYSYDRKTKRPSHKGAGAKGENIFQVFLVKDASTOLANN 425
QY 341 QY 342
Db 426 TY 427

RESULT 5
EX02.BPT4
ID EX02.BPT4 STANDARD; PRT; 560 AA.
AC P04522;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EXONUCLEASE SUBUNIT 2 (EC 3.1.11.-) (PROTEIN GP46).
GN 46.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
OC T4-like phages.
OX NCBI_TaxID=10665;
RN [1]
RX MEDLINE=85257446; PubMed=4018026;
RA Gram H.; Rueger W.;
RT "Genes 55, alpha q1, 47 and 46 of bacteriophage T4: the genomic
RT organization as deduced by sequence analysis.";
RL EMBO J. 4:257-264 (1985).
RN [2]
RX SEQUENCE FROM N.A.
RA Kutter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;
RT "Bacteriophage T4 genome analysis.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RX SEQUENCE OF 478-560 FROM N.A.
RA Hsu T., Karam J.;
RL Submitted (APR-1987) to the EMBL/GenBank/DBJ databases.
CC 1- FUNCTION: EXONUCLEASE INVOLVED IN PHAGE DNA RECOMBINATION,
CC REPLICATION, AND REPAIR.
CC 1- SUBUNIT: CONSISTS OF TWO SUBUNITS: GP47 AND GP46.
CC 1- SIMILARITY: STRONG TO T5 PROTEIN D13 AND TO YEAST RAD52.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X01804; CAA25945.1;
CC EMBL; M15080; AAA32516.1;
CC EMBL; AF158101; AAD42472.1;
CC EMBL; M10160; AAC05392.1;
CC PIR; A04298; NCBP6.
CC NP_BIND 36 43 ATP (POTENTIAL).
SQ SEQUENCE 560 AA; 63613 MW; D107829431BAB2FC CRC64;

Query Match 3.7%; Score 122; DB 1; Length 560;
Best Local Similarity 18.7%; Pred. No. 0.16;
Matches 102; Conservative 78; Mismatches 168; Indels 198; Gaps 22;

QY 65 NYEALPIYKNGDIFVRHLKIOVKYKETOGIT-----SSGFASLTFTGTLGAPIIPR 117
Db 3 NFKLRVKYKINMSVGONGIDQLDRVOKTLITGRNGGKSTMLEAITP-GLFGKPRDV 61

QY 118 TSSRYFNFTEDHKMVEALRWASTHMSPTLLKLCDDVQPMQY-----F 152
Db 62 KKGQLINSTNKKELLVE--LW-----MEYDEKKYIKRGQKPNVF 99

QY 163 DLTC-----QLLGKAEVDGASFLKLVWDGTR--TPF-----PSWRV 196

```

```

Db 100 EITVNGTRLNESASSKDFQAEFEQLIG---MSYASFQKQIVLGTAGYTPFMGLSTPARR 156
QY 197 LIQDLVLEGDLSHIHRQLNLTIDILVYDNHVVHVARSLKVGSEFLRYSLSLTKLQ-SMNSEN 255
Db 157 LVEDLLEVTGLAEMDKLNKALIRELNSQVLDVKKDSIIQIKIYNDNVERQKLTGDN 216
QY 256 OTMLSLEFHLHGTSYGRIRVLPSNSDSDVDKKDLKSANUTANOHSVDVICQSEPDPSF 315
Db 217 LTRLQ-----NMYDOLAKEARTLKSEIEAN---ERLVNVLDEPTDAF 258
QY 316 PSSGS-----VSLY-----EVERC-----OO 331
Db 259 NKIGQEAFLIKSKIDSYNKVINNYHEGGLCTCLSQLSGDKVVKSKDKVSECTHSFEQ 318
QY 332 LSA-----TILTHQYLERTPLCAILK-----OKAPOQYRIR 363
Db 319 LSTHRDLKVLVDEYRDNIKTQOSLANDIRNKKQSLIAAVDKAKKAAIEKASSEFIDH 378
QY 364 AKLRSYKPRRLFSQVKLHCPKCHLLOEVPHEGDLIIIFQDGATKTPVVKLQNTSLYDSKI 423
Db 379 ADEIALLOEELDKIVK---TKTNLVMEKYHKGILTDMKDSGKIGAIK-KYIPLFNKOI 434
QY 424 WTKKQKGRKVAHFVKNGNGLPLSNCELLIEGGTLSEICKLSNKFNSVIPRSGHEDL 483
Db 435 -----NHYLKIMEA---DIVFTLDEEFNETIKSR-GREDF 465
QY 484 ELLDLS 489
Db 466 SYASFS 471

RESULT 6
GATL_YEAST STANDARD; PRT; 510 AA.
AC P43574;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSCRIPTIONAL REGULATORY PROTEIN GATL.
GN GATL OR YFL021W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GRF88;
RX MEDLINE=96182087; PubMed=8622886;
RA Coffman J.A., Rai R., Cunningham T., Svetlov V., Cooper T.G.;
RT "Gatlp, a GATA family protein whose production is sensitive to
RT nitrogen catabolite repression, participates in transcriptional
RT activation of nitrogen-catabolic genes in Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 16:847-858 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268 (1995).
CC 1- FUNCTION: POSITIVE REGULATOR OF MULTIPLE NITROGEN CATABOLIC GENES.
CC 1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC 1- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC

```

```

Db 69 VNVFFSQNFEDLPIIORVGDIYVRHARLQHYNDARLQNVNMYRSMCLFIGNDKEAP 128
Qy 114 IIPRT-----SSRYF-----NFTTEDH--KMVEALRVWASTHMSPSWTL--LKLC 154
Db 129 LEFVENEEDTNYFSTPPNFSGKSFQTOGHETKILKDLKWSKDYFSNNDVVEQVKA 188
Qy 155 DVOPMOY-----FDLTCLLGAKEVDSGASFLKLVMDGT-----RTFFPSWEVLIQDL 201
Db 189 DIETAMKNKTFDOLLAKRVTEISNDQVNTVSLNDSTGQTWTGHLFRKRFPH----- 240
Qy 202 VLEGDLSHIHLQNLITDILVYDNHVVHVARSLKVGSLFRLIYSWHLKLSNMSEN 255
Db 241 LVKGDVLRKSVAKEDNSLIFSSHNLK-----FFSFSSIHKKLKSSISSU 288

RESULT 2
TEBH_EUPCR
ID TEBH_EUPCR STANDARD; PRT; 460 AA.
AC Q06183; 1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 01-JUN-1994 (Rel. 29, Last annotation update)
DE TELOMERE-BINDING PROTEIN HOMOLOG.
OS Euplotes crassus.
OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotida; Euplotidae;
OC Euplotes.
OX NCBI_TaxID=5936;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93126105; PubMed=1480483;
RA Wang W., Skopp R., Scofield M., Price C.;
RT "Euplotes crassus has genes encoding telomere-binding proteins and
telomere-binding protein homologs."
RL Nucleic Acids Res. 20:6621-6629(1992).
CC -1- FUNCTION: MAY BIND TELOMERIC T4G4 SEQUENCES.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT
FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EXBL: M96819; AAA29128.1; -.
DR PIR: S35525; S35525.
DR InterPro: IPR003415; Telo_bind_alpha.
DR Pfam: PF02307; Telo_bind_alpha; 1.
DR DNA-binding; Nuclear protein; Telomere; Multigene family.
KW SEQUENCE 460 AA; 53360 MW; ED1C141385A0B5FE CRC64;

Query Match 4.98; Score 163.5; DB 1; Length 450;
Best Local Similarity 21.66; Pred. No. 0.0001.
Matches 117; Conservative 74; Mismatches 195; Indels 155; Gaps 24;

Qy 9 YIYTPNLQ--KGGTIVNVYGVYKFFKPPYLSKGTGYCSVVIIVDOT-NVK-----LTC 59
Db 19 YEYTEIGSEIEENEASINFYAVVIDACEFYKVKDEKMYCYLKVITDITHNVKGGDUNFAIV 78
Qy 60 LIFSGNTEALPIIYKNGDIVRHLRLKIOYKKEIOGIT-----SSGFASLTFEGTLGAPI 114
Db 79 ALQSRKFDELPIIQCGDIIIRVHRAEYN-YKDDQHYFKLNYSYSSWALFSADEEVAPEV 137
Qy 115 IP-----RTSSKYVFTTEDHKMVEALRVWASTHMSPSWTLKLCQVCPMOYFD 163
Db 138 KDEGDDFTYRSYAYSGQYNEFDQDQKLLNTRANNKSYFAKN-----DV----- 183
Qy 164 LTCQLLGRAEYDVGASFLKLVMDGTTRTPPSWRVLIDLVLEGDLSHIHLQNLITDILVY 223

```

```

Db 184 -----IIDEMYTPLSQAROEGBDFNVGVKVTQIVHRDYITSDLRVK 224
Qy 224 DN-----HVHVARSLKVGSLFRLIY-SLHTKLSQMSNSENQTMLSLEPHLHGGSYGRGIRV 277
Db 225 DISKATWELTVSRR-----KFRPLYEGVLIKIRSVNIDSE-----TERERCLEL 268
Qy 278 LPESN-----SDVDLKKDLKLESANLTANQHS-----VICQSEPDSDSPSSGVSLEYEVR 328
Db 269 AHSNIMTFVPFSRLAKSLDSQISLSPDKVLELIKVKVILTEP----- 311
Qy 329 COOLSATILTDHOYLERTPLCAILKOKAPQ--YRIRAKLRYSYKPRRLQFQSVKLHCPKCH 386
Db 312 --VLAITTFGDIYSELPLTELSEIFEVDVTDKDAVFRARFSEILKITPDVDEYVEEYTPK-- 367
Qy 387 LLOEVPHEGDDLIIFODGATKT--PVVKLO-----NTSLYDS--KIWT-TKNOKGRKVAV 436
Db 368 -----GAPRSKPYKVQFLIKDPSTALNDNLKIYLYSHGDLGKEFF- 409
Qy 437 HFVKNNGIILPLSNECLLLIEGGTLSEICKLSNKFNSVIPVRSGHEDLELDLSAPFLIOG 496
Db 410 -----PGVDPSAQ-----TPSGHSLK-RKYASTLKMKNVHIDAVLEKVGGAFFIRD 455
Qy 497 T 497
Db 456 T 456

RESULT 3
TEBA_OXYNO
ID TEBA_OXYNO STANDARD; PRT; 495 AA.
AC P29549;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT (TELOMERE-BINDING PROTEIN 56
DE KDA SUBUNIT) (TEBP ALPHA).
GN MAC-56A AND MAC-56K AND MAC-56S.
OS Oxytricha nova.
OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Stichotrichida;
OC Oxytrichidae; Oxytricha.
OX NCBI_TaxID=5945;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=92035001; PubMed=1840510;
RA Gray J.T., Celandier D.W., Price C.M., Cech T.R.;
RT "Cloning and expression of genes for the Oxytricha telomere-binding
RT protein: specific subunit interactions in the telomeric complex."
RL Cell 67:807-814(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92249771; PubMed=1577273;
RA Mitcham J.L., Lynn A.J., Prescott D.M.;
RT "Analysis of a scrambled gene: the gene encoding
RT alpha-telomere-binding protein in Oxytricha nova."
RL Genes Dev. 6:788-800(1992).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=99091054; PubMed=9875850;
RA Horvath M.P., Schweiker V.L., Bevilacqua J.M., Ruggles J.A.,
RA Schultz S.C.;
RT "Crystal structure of the Oxytricha nova telomere end binding protein
RT complexed with single strand DNA."
RL Cell 95:963-974(1998).
CC -1- FUNCTION: MAY FUNCTION AS PROTECTIVE CAPPING OF THE SINGLE-
CC STRANDED TELOMERIC OVERHANG. MAY ALSO PARTICIPATE IN TELOMERE
CC LENGTH REGULATION DURING DNA REPLICATION. BIND SPECIFICALLY TO THE
CC T4G4-CONTAINING EXTENSION ON THE 3' STRAND AND PROTECT THIS REGION
CC OF THE TELOMERE FROM NUCLEASE DIGESTION AND CHEMICAL MODIFICATION.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- MISCELLANEOUS: THE SEQUENCE OF THE A (OR ALANINE) VERSION IS
CC SHOWN. THE S (OR SERINE) VERSION DIFFERS IN ONLY TWO POSITIONS,
CC AND THE K VERSION IN ONLY ONE. THERE MAY BE OTHER VERSIONS IN
CC

```

C;Genetics:
A;Gene: CES
A;Map posit
A;Introns:

Query Match 3.5%; Score 117.5; DB 2; Length 251;
Best Local Similarity 24.5%; Pred. NO. 0.23;
Matches 46; Conservative 28; Mismatches 65; Indels 49

Qy	160	OYFDLTCOLLGKAEVDCGSFLLKWDG-----TRTPFSHRV----	196
Db	25	RYFDVLAQHVSVYETINGAOMLRVGRKFGPGCASERBRLFHVTDQSFXYIVPPNP	84
Qy	197	LIQDLVLEGDLSHIRLONLTIDILYVDNHVHVARSCLKVGSFLRIYVSLHTKLOSMNSEQ	256
Db	85	RIGKAIEENG-----KELLIEIDYDHRGLKMLNSGDFVATQNVHAA-----STRO	132
Qy	257	TMLSLFHLH--GGTSYSGRGRIVLPE--SNSVDQLKKDLESA-----NLTAQHSDV	305
Db	133	TEMQV---LHGGSASYRGITTPVDFEHEAFONFKKKVAVLTVAYDENTETEQOPEE	189
Qy	306	ICQSEPPD	313
Db	190	VAENHVDE	197

RESULT	12
1. The results of the study show that the proposed method is effective in reducing the number of false alarms and improving the detection rate of the system.	
2. The proposed method is able to detect and classify the different types of faults in the system.	
3. The proposed method is able to detect and classify the different types of faults in the system.	
4. The proposed method is able to detect and classify the different types of faults in the system.	
5. The proposed method is able to detect and classify the different types of faults in the system.	
6. The proposed method is able to detect and classify the different types of faults in the system.	
7. The proposed method is able to detect and classify the different types of faults in the system.	
8. The proposed method is able to detect and classify the different types of faults in the system.	
9. The proposed method is able to detect and classify the different types of faults in the system.	
10. The proposed method is able to detect and classify the different types of faults in the system.	
11. The proposed method is able to detect and classify the different types of faults in the system.	
12. The proposed method is able to detect and classify the different types of faults in the system.	

T00342
hypothetical protein KIAA0580 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00342
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14086; MUID:98290545
A:Accession: T00342
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1044 <NAG>
A:Cross-references: EMBL:AB011152; NID:g3043683; PIDN:BAA25506.1; PID:g3043684
A:Experimental source: brain; clone HJ0601
C:Genetics:
A:Note: KIAA0580

Query Match 3.5%; Score 116; DB 2; Length 1044;
Best Local Similarity 22.0%; Pred. NO. 2.6;
Matches 125; Conservative 83; Mismatches 212; Indels 148; Gaps 33;

Oy	49	I V D O T N Y K L T C L L --- F S G N Y E A L P I Y K N G D I V R H R L K I O V Y K K E T O G I T S S G F A S L T	105
D8	255	L L E E T N - K W C V L E G G F S Y E N D K S T P N G T I N I N E V I A I H K B E D F Y L N T G P I F --- I	310
Oy	106	F E G T L G A P I P T R S S K Y F N E T T I E D H K M V E A L R Y W A --- S T H A S P S W T L L K L C O V O P M O Y F	162
D8	311	F E ----- I V L P S E R V F E G A E T S --- Q A O R K W E T A I A K H F V P L F A - E N L T E A D --- Y	355
Oy	163	D L T C O L L G K A E V D G A S F E L L K V W --- D G T R T P P S W R V L I O D L V L E G D L S H I R L O N L T I D	219
D8	356	D L I Q L F Y K - D C H A L D O W R K G W A M D K S S L I F --- C L O M E O V O G D R M H L R L Q E L I T S	409
Oy	220	I L V D T N H - V H V A R S L K V G S F E L R I Y S L H T K L ----- O S M S E N O T	257
D8	410	T M V O N G E K L D V L L V E R G R T L Y I H G - H T K L D F T V W H T A I E K A G T O G N A L Q O O L S K N D V	468
Oy	258	M L S F E F L H U G G T S Y G R G I R V L P E S N S D ----- V D O L K K D L S A N L T A N Q H - S D V I C -	307
D8	469	P I I V N S C I A F V T Q Y G L A C K V I Y O K N G D P L H I S E L L E S F K K D A R S F L R A G K H O L E D V A T	528
Oy	308	--- Q S E P D D S F P S S G S V S L Y ----- E V E R C O O L S A T I L D H O Y L E R T P L C A I L	352

```

Db      529  LKSFSLDIDALLTK---ELYPWISALTDODDKERIKKYGA-FIRSLFCVNPRTALAAII 581
Qy      353  KQAKPOOYRIRA-----KLASYKPRRLFQGVKLUHCPKCHLLQZVPHEGDL-----DIIFOD 403
Db      585  E-----HLRYQKRCSEINHMNAHLALVFSSC-LFQTKGOTSEEVNYIEDLINNYVEIFEV 639
Qy      404  GATKTPVVKLQNTSLYDSKIW--TTKNQKGRKVAVHFVKNNGILPLSNECLLIEGGTILS 461
Db      640  KEDQVKQMDIENSFIHK---WKDTQVSOAGDLLIEVTVERK-----EPDCSIIIRISPMV 691
Qy      462  EICKLSN---KFNSVIPVRSG-----HEDE-----LLD-----LSAP--- 491
Db      692  EAEELTNDILAIKNIPITKGDWATTFEVIENEELERPLHYKENVOLRWSSLAEPGSA 751
Qy      492  -----FLIOGTTHYGVCKOCSSLSRSIQ 513
Db      752  YLVVKRELTADTIKH--CDSRSTLGSIK 777

RESULT 13
S56233
probable membrane protein YFL021w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 06-Feb-1999
C:Accession: S56233
R:Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma
submitted to the EMBL Data Library, May 1995
A:Description: Analysis of the nucleotide sequence of chromosome VI from S
A:Reference number: S56186
A:Accession: S56233
A:Molecule type: DNA
A:Residues: 1-510 <MUR>
A:Cross-references: EMBL:D50617; NID:g836685; PID:d1009858; PID:g836733; M
C:Genetics:
A:Gene: SGD:GAT1
A:Cross-references: SGD:S0001873; MIPS:YFL021w
A:Map position: 6L
C:Superfamily: GARA-type zinc finger homology
C:Keywords: transmembrane protein; zinc finger
F:14-30/Domain: transmembrane #status predicted <TM1>
F:307-360/Domain: GARA-type zinc finger homology <GF>

```

Query Match 3.5%; Score 115.5; DB 2; Length 510;
Best Local Similarity 17.6%; Pred. No. 0.95;
Matches 86: Conservative 59; Mismatches 132; Indels 213; Gaps 17;

Qy	209	HIHRLQNLITDILVDYDNRV - HVARSKVGSEFLRIYSUHTKLOKSHNSNQTMLSLEFHJHG	267
Db	27	YIHTCTIVVWITMSTNRVNPDLNLK - EIWDLYSSAQKILPDSNRILNLSWRLHN	84
Qy	268	GTSYGRGIRVLPESNDVD -	286
Db	85	RISFHRINRIHOHNSIMDFSASFAGVGNNAAGPGNNDDLTDTDNOQPFLLSDMNLNGSS	1444
Qy	287	-----OLKKDLESANLTANOHS-----DVICQSEPPDDSF-	315
Db	145	VFENVEDDDDDDDVETHSIVHSDLLNDMSAQSRASHNASGPFNFLTSCSSFFDDHFI	2040
Qy	316	-----PSSGSVSLYEVERCQ---QLSATILTDHOYLERPLCA	350
Db	205	FTNNLPFLNNNSINNNHSHNSHHNNGPSIANNTNANTNTNTSASTNTNPGLLRNPS	2640
Qy	351	ILKOKAPOQYRIRAKLRYSKPRRLFQSVYKLHCPRCHLLQVEPHGEDLDIIFQDQAT----	4060
Db	265	IVKPCSRNRSSVRKKKPAKKIKSSTS-	3010
Qy	407	--KTPVVKLQNTSLYDSKIWTTKQKGRKAVHEVKNNGITPLSNEC--LLLLEGGT--L	4600
Db	302	TSSNPDIKCSNCTTSTTPLW-RKDPKG-----LPLCNACGLFKLHGVTPL	3470
Qy	461	S-----EICKLSKNKFSNVIPVR	4770

A:Experimental source: clone F39H2

C:Genetics:

A:Gene: CESP:F39H2.1

A:Map position: 1

A:Introns: 42/3; 104/2; 160/3; 273/2; 317/2; 429/2; 502/2

Query Match 3.9%; Score 130.5; DB 2; Length 587;
Best Local Similarity 18.9%; Pred. No. 0.089;
Matches 109; Conservative 72; Mismatches 162; Indels 233; Gaps 29;

QY 122 YNFTTDEHKWVE-----ALRVWASTHMSPSWTLKLLKCDVQPMQYFDLT-----CQL-- 168
DB 71 YFDVLAQVSHVYVETKNGSWTLRVWRAQFGPEASKE-----REMLFHVITNTKCYIVP 126
QY 169 ----LGKA-EVDGASFLKLVMDGTRTPFPSPMRVLIDLVLEGLDLSHRLQNLITIDILVY 223
DB 127 PPRIGKAVEESGKEFL-----LEIDVY 149
QY 224 DNHVHVARSLKYGSLRIYSLH-----TKLSMSENQTMLSLEFHLHGTSYGRGIRV 277
DB 150 DDBRIGIEKLSGDFVALQNVHAASVGLTEMQVLHG-----GGOAYNRGISK 196
QY 278 LPES--NSVDQKLDLESA-----NLTAHQSDVICQSEPDSPSPSGSVSLYEVE 327
DB 197 VPVDFRNEAFQIFKRVESVLEAVTDCDNFIEFOKENVVOQSETSTQTSSTNSGDADLK 256
QY 328 RCQOLSA-----TILT-----DH-----QYLERIPLCAILKOKAPOQYR 361
DB 257 HCQIKIGEDISVDYFLKKSGRYNFLTHAHDHYRGLDKWTRSYVC-----SPETAK 310
QY 362 IRALKRSY-----KPRLPFSQVHLKCPKCHLLOEVPHE-----GDLDIIF 401
DB 311 LLPHIMGYTAGDPPAGLNPJLE-----ENVPKRFDSQVTLVNNHCPGAVMEVF 361
QY 402 Q-----DGATKTPVVKLQNTSLYDSKIWTIKQKGRKVAHVFKNNGI--- 444
DB 362 EGSKIEIAGGAVLCGTGDFRADKMFLESUKPGNLHWTE-----IKFGIYLDNTYFSLD 417
QY 445 LPLSNPC-----LLLEGGTLSEICKLSNKFNSVIPV-RSGHEDL----- 483
DB 418 MPFPERCEAEKILL-----KAIEAPHENIVIPUHLRGRELQAIIRLNEPIWYD 470
QY 484 -----ELDLSAPFLIOGTHHYGCKQCSSURSTONLSNLDKTSWIP-SSVAEALGIV 536
DB 471 ERKVISDLDLDFSS-----NGVLKQYRDI-KVIKKGSYCEISONSVIIDIS 515
QY 537 PLOTYVFWTFITLDDGTCVLEYALYLDSDKFFOIPASE 572
DB 516 MLXYTF-----GNGV-----NDDEGIIRIPYSD 538

RESULT 8
NCBPX6
exonuclease 46 (EC 3.1.1.1.-) chain 2 - phase T4
C:Species: phase T4
A:Note: host Escherichia coli
C:Date: 17-Mar-1987 #sequence_revision 21-Nov-1997 #text_change 19-Jan-2001
C:Accession: A04298; T10162
R:Gram, H.; Ruger, W.
EMBO J. 4, 257-264, 1985
A:Title: Genes 55, alpha-qt, 47 and 46 of bacteriophage T4: the genomic organization as
A:Reference number: A91016; MUID:85257446
A:Molecule type: DNA
A:Accession: A04298
A:Residues: 1-560 <GRA>
A:Cross-references: GB:X01804; NID:gl15229; PIDN:CAA25945.1; PID:g577852
A:Note: the authors translated the initiation codon GUG for residue 1 as Val
C:Genetics:
A:Gene: 46
A:Start codon: GUG
C:Superfamily: phase T4 exonuclease 46
C:Keywords: exonuclease; hydrolase; nucleotide binding; P-loop

F:36-43/Region: nucleotide-binding motif A (P-loop)

Query Match 3.7%; Score 122; DB 1; Length 560;
Best Local Similarity 18.7%; Pred. No. 0.36;
Matches 102; Conservative 78; Mismatches 168; Indels 198; Gaps 22;

QY 65 NYEALPIYKNGDIVRPHRLKIQVYKKEGKIT-----SSGFASLTFFEGTIGAPIR 117
DB 3 NFKLNRVYKYNIMSVGQNGDIQLDKVQKTLITGRNGGGKSTMLEAITF-GLFGKPFQV 61
QY 118 ISSYVENFTTDEHKMVALRWASTHMSPSWTLKLLKCDVQPMQY-----F 162
DB 62 KKGOLINSTNKKELLE-LW-----MEYDEKYYIKRGQKPNVF 99
QY 163 DLTC-----OLLKAEVDGASFLKLVMDGTR--TPF-----PSWRV 196
DB 100 EITWGTIRLNESASSKDFQAEFEQLIG---MSVASFQIIVLGTAGTPTFPMGLSTPARRK 156
QY 197 LIQDILVLEGOLSHIRLQNLITIDILVYDNHVVHVARSLKVGSLRIYSLHTKLO-SMSEN 255
DB 157 LVEDLLEVTGLAEMDKLNLKALIRELNSONQVLDVKKDSIIQOIKIYNDNVEROKKLTGDM 216
QY 256 QTMLSLEFHLHGTSYGRGIRVLPESNDVDQKLDLESANLTANQHSDDVICQSEPDSP 315
DB 217 LTRQ-----NMYYDLAKEARTLAKSETEAN---ERLVNIVLDEDDPTDAF 258
QY 316 PSSGS-----VSLY-----EVERC-----QQ 331
DB 259 NKIGQAEFLIKSIDSYNKVINMYHEGCLPTCLISQSSGDKVSVSKDKVSECTHSEFQ 318
QY 332 LSA-----TILTDHOYLERTPLCAILK-----OKAPOQYRIR 363
DB 319 LSTHRDLNLKVLVDEYRDNKIQOQLANDIRNKKQSLIAAVDKAKKAAIEKASSEFIDH 378
QY 364 AKLSYKPRRPFQSVKHLCPKCHLLOEVPHEGDDLIIFODGATKTPVVKLQNTSLYDSKI 423
DB 379 ADEIALLOEELDKIVK---TKTNLVMEKYHRRGILTDMKDSGIGKAIK-KYIPLFNKQI 434
QY 424 WTIKRNKGRKVAHVFKNNGILPLSNECLLIEGGTILSEICKLSNKFNSVIPVRSGHEDL 483
DB 435 -----NHYLKIMEA---DYVFTLDEEFNETIKSR-GREDF 465
QY 484 ELLDLS 489
DB 466 SYASFS 471

RESULT 9
A54810
TMV resistance protein N - tobacco (Nicotiana glutinosa)
C:Species: Nicotiana glutinosa
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 05-Nov-1999
C:Accession: A54810
R:Whitham, S.; Dinesh-Kumar, S.P.; Choi, D.; Hehl, R.; Corr, C.; Baker, B.
Cell 78, 1101-1115, 1994
A:Title: The product of the tobacco mosaic virus resistance gene N: similarity to tol
A:Reference number: A54810; MUID:95007759
A:Accession: A54810
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1144 <WHI>
A:Cross-references: GB:U15605; NID:g558886; PIDN:AAA50763.1; PID:g558887
C:Genetics:
A:Gene: N
A:Introns: 160/2; 525/3; 616/3 1139/3
Query Match 3.7%; Score 122; DB 2; Length 1144;
Best Local Similarity 18.2%; Pred. No. 1.1;
Matches 135; Conservative 106; Mismatches 200; Indels 300; Gaps 36;

QY 71 IIVKNGDIVRFHRL-----KIQVYKKEGKIT-ITSSGFASLTFFEGTIGAPIRPTSSKYPN 124

C: Superfamily: Schizosaccharomycetes hypothetical protein SPAC26H5.06

[illegible]

Ov 437

437 HEVKNNGIIPLSNCEIIJJEGGTLSEICKLSNKFNSVTPVRSCHDELELLDLSAPE

QY 57

CC inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAH77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SQ Sequence 675 BP; 213 A; 103 C; 128 G; 229 T; 2 other;

Query Match 10.6%; Score 201; DB 22; Length 675;
 Best Local Similarity 100.0%; Pred. No. 1.2e-45;
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1705 ccagcatcagaagttctgatgatgatcccttcagaaagtgtgatatacatcatgat 1764
 Db 1 ccaagcatcagaagttctgatgatgatcccttcagaaagtgtgatatacatcatgat 60
 Qy 1765 atgtttgtccctccaggaataaaattgatcatatccgtgttggaatgcttcacaa 1824
 Db 61 atgtttgtccctccaggaataaaattgatcatatccgtgttggaatgcttcacaa 120
 Qy 1825 tcatacaatgtcacaaatggacagacagataatacaatttgcatactcagattttgacaccaca 1884
 Db 121 tcatacaatgtcacaaatggacagacagataatacaatttgcatactcagattttgacaccaca 180
 Qy 1885 gttgcagaagatgtaattctaa 1905
 Db 181 gttgcagaagatgtaattctaa 201

RESULT 12
 AAI39284
 ID AAI39284 standard; DNA; 475 BP.
 XX
 AC AAI39284;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #7970 used to measure gene expression in human placenta sample.
 XX
 KW Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX

PS Claim 25; SEQ ID No 7970; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 475 BP; 191 A; 70 C; 69 G; 145 T; 0 other;

Query Match 7.5%; Score 143.6; DB 22; Length 475;
 Best Local Similarity 91.6%; Pred. No. 7.4e-30;
 Matches 152; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1140 tcattgccctaaatgtcatttctgccaagaagttccacatgagggcgatttggataata 1199
 Db 310 tcattcattgcccaataattttaggcaagaagttccacatgagggcgatttggataata 369
 Qy 1200 ttctcaggatggtgcaactaaaccccaagttgcaagttacaataacatcattatga 1259
 Db 370 ttctcaggatggtgcaactaaaccccaagttgcaagttacaataacatcattatga 429
 Qy 1260 ttcaaaaatctggaccactaaaatcaaaaaggacgaaaaagtagca 1305
 Db 430 ttcaaaaatctggaccactaaaatcaaaaaggacgaaaaagtagca 475

RESULT 13
 AAF58252/C
 ID AAF58252 standard; DNA; 936 BP.
 XX
 AC AAF58252;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Oligonucleotide D1835.
 XX
 KW Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 XX
 OS Synthetic.
 XX
 PN WO200107665-A2.
 XX
 PD 01-FEB-2001.
 XX
 PF 26-JUL-2000; 2000WO-US20476.
 XX
 PR 26-JUL-1999; 99US-0145695.
 PR 17-MAR-2000; 2000US-0190259.
 XX
 PA (CLIN-) CLINICAL MICRO SENSORS INC.
 XX
 PI Umek RM;
 XX
 DR WPI; 2001-159728/16.
 XX
 PT Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface -
 XX
 PS Example 6; Page 127; 159pp; English.
 XX
 CC The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETW) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 XX
 SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification; the primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 568 BP; 167 A; 115 C; 96 G; 178 T; 12 other;

Query Match 22.0%; Score 419; DB 22; Length 568;

Best Local Similarity 93.2%; Pred. No. 6.1e-106;

Matches 465; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 1411 aatagttaattctctgtgagatcgccacgaag--acctggaacttttggacctttcag 1468

DB 541 ANAAGGGAATTCGGGGAATNGGCCAAGGAGGACCTGGGAATTTGGACCTTTTCAG 482

QY 1469 caccattttttacagaagaacaatacatcactatggtatgaacagtggttttagtga 1528

DB 481 CACCATTTTATACAGGGAACAAAACATCACTAGGATGTAACAGTGTNTAGTTGA 422

QY 1529 gatccatacaaatctaaatt--ccctggttgataaacaatcgatgattccttc-ttctgt 1586

DB 421 GATCCATACAAAATNTAAATTCCTCGTGTGATAAACAICGTGGATTCNTTCTTCGT 362

QY 1587 ggcagaagcactgggtattgtaccctccaatattgtttgttatgacctttacacttga 1646

DB 361 GGCAGAGCAGCTGGGTATTGNACCCCTCCAATATGTTGTATGACCTTTACACTTGA 302

QY 1647 tgatgaacagagactactagaagcctatctctatggtattctgacaaaattcttcagattcc 1706

DB 301 TGATGAACAGAGGACTAGAGAGCCTATNTCATGATTNTGACAAATNTTCCAGATTCC 242

QY 1707 agcatcagaagttctgagatgagcttcacgttcagaaaagtatgatcatgatgat 1766

DB 241 ACATCAGAGGTTCTGTGATGATGACCTTCAGAAAAGTGTGGATATGATCATGGATAT 182

QY 1767 gttttgtctccaggaataaaaattgatgcataatccgtggttggaatgcttcatcaagtc 1826

DB 181 GTTTGTCTCCAGGAATAAAAATTGATGCATATCCGTGGTTGGAATGCTTCATCAAGTC 122

QY 1827 atacaatgtcaaatggaacagataatacaatttgcattatcagatttttgacaccacagt 1886

DB 121 ATACAATGTCAAAATGGAACAGATAAATCAAAATTTGCTATCAGATTTTGTACACACAGT 62

QY 1837 tgcagaagatgtaattctaa 1905

DB 61 TGCAGAAGATGTAATCTAA 43

RESULT 9

AAH05347

ID AAB05347 standard; cDNA; 866 BP.

XX AC AAH05347;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:2182.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

KW

XX Homo sapiens.

OS EP1074617-A2.

PN 07-FEB-2001.

PD 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

DR

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs

XX Claim 1: SEQ ID 2182; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX Sequence 866 BP; 254 A; 162 C; 171 G; 274 T; 5 other;

SQ

Query Match 18.0%; Score 342; DB 22; Length 866;

Best Local Similarity 91.2%; Pred. No. 1.4e-84;

Matches 384; Conservative 0; Mismatches 30; Indels 7; Gaps 2;

QY 1 atgtcttgggtccagcaacaaattatatatacacccctggaatcaacttaagggtggt 60

DB 439 atgtcttgggtccagcaacaaattatatatacacccctggaatcaacttaagggtggt 498

QY 61 acaattgtcaatgtctatgtgtgtgaagttctttaagcccccatctatcaagcaaaagga 120

DB 499 acaattgtcaatgtctatgtgtgtgaagttctttaagcccccatctatcaagcaaaagga 558

QY 121 actgattattgtcagttgttaactattgtggaccagacaaatgataaacttaacttgcctg 180

DB 559 actgattattgtcagttgttaactattgtggaccagacaaatgataaacttaacttgcctg 618

QY 181 ctcttttagtgaactatgaagccctcccaataattataaaaatggagatattgttcgcg 240

XX

```
Oy 1580 cttctgtgagcagaactgggtattgtaccctcccaatatgtgtttgttatgaccttta 1639
Dy 601 cttctgtgagcagaactgggtattgtaccctcccaatatgtgtttgttatgaccttta 660
Oy 1640 caattgatgatgaacagagactagacagcctatctcatgattctcaaaaattcttcc 1699
Dy 661 caattgatgatgaacagagactagacagcctatctcatgattctcaaaaattcttcc 720
Oy 1700 agattccagcatcagaagtctctgattgattgacattcagaaaagtgtggatgatca 1759
Dy 721 agattccagcatcagaagtctctgattgattgacattcagaaaagtgtggatgatca 780
Oy 1760 tggatattgtttgtctccagagaataaaaaattgatgcataatcgtgtgtggaatcttca 1819
Dy 781 tggatattgtttgtctccagagaataaaaaattgatgcataatcgtgtgtggaatcttca 840
Oy 1820 tcaagtcatacaatgtccaaatgaacagagataatcaaaattgtctatcagatttttgaca 1879
Dy 841 tcaagtcatacaatgtccaaatgaacagagataatcaaaattgtctatcagatttttgaca 900
Oy 1880 ccacagttgcagaagatgtaattctaa 1905
Dy 901 ccacagttgcagaagatgtaattctaa 926

RESULT 6
AAH05303
ID AAH05303 standard; cDNA: 576 BP.
XX
AC AAH05303;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:2138.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 2138; 2537pp + CD ROM; English.
XX
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
```

```
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 576 BP; 167 A; 113 C; 114 G; 173 T; 9 other;

Query Match 28.6%; Score 545; DB 22; Length 576;
Best Local Similarity 98.7%; Pred. No. 9.1e-141;
Matches 545; Conservative 0; Mismatches 7; Indels 0; Gaps 0:

Oy 1 atgtctttgttccagcaacaatttatatatatacacccctgaatacaacttaagggtggt 60
Dy 25 atgtctttgttccagcaacaatttatatatatacacccctgaatacaacttaagggtggt 84
Oy 61 acaattgtcaatgtctatgtgtgtgaagtcttttaagccccccatataatgaagcaagga 120
Dy 85 acaattgtcaatgtctatgtgtgtgaagtcttttaagccccccatataatgaagcaagga 144
Oy 121 actgattattgtcagttgtaactattgtgaccagacaaaatgtaaaactaacttgcctg 180
Dy 145 actgattattgtcagttgtaactattgtgaccagacaaaatgtaaaactaacttgcctg 204
Oy 181 ctcttagtggaactatgaagcccttcccaataattataaaatggagatatgttcgc 240
Dy 205 ctcttagtggaactatgaagcccttcccaataattataaaatggagatatgttcgc 264
Oy 241 ttccacaggtgaagattcaagtataataaaaggagaccaggtatcaccagctctg 300
Dy 265 ttccacaggtgaagattcaagtataataaaaggagaccaggtatcaccagctctg 324
Oy 301 ttgcatctttgacgtttgagggaaactttggagccccctatcatcctgcacttcaagc 360
Dy 325 ttgcatctttgacgtttgagggaaactttggagccccctatcatcctgcacttcaagc 384
Oy 361 aagatttttaactcactactgagaccacaaaatgtagaagccttaagtgtttggcca 420
Dy 385 aagatttttaactcactactgagaccacaaaatgtagaagccttaagtgtttggcca 444
Oy 421 tctactcatatgtcaccgtcttggacattactaaaattgtgtgattcagccaatgcag 480
Dy 445 tctactcatatgtcaccgtcttggacattactaaaattgtgtgattcagccaatgcag 504
Oy 481 tattttgacctgactgtcagctcttgggcaagcaagagtgagcagcattctttt 540
Dy 505 tattttgacctgactgtcagctcttgggcaagcaagagtgagcagcattctttt 564
Oy 541 ctaaaaggtatgg 552
Dy 565 ctaaaaggtatgg 576

RESULT 7
AAH07649
ID AAH07649 standard; cDNA: 854 BP.
XX
AC AAH07649;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:4484.
XX
```

26-JUL-2001.
26-DEC-2000; 2000WO-US34263.
21-JAN-2000; 2000US-0488725.
25-APR-2000; 2000US-0523317.
09-JUL-2000; 2000US-0598042.
19-JUL-2000; 2000US-0620312.
03-AUG-2000; 2000US-0653450.
14-SEP-2000; 2000US-0662191.
19-OCT-2000; 2000US-0693036.
29-NOV-2000; 2000US-0727344.
(HYSE-) HYSEQ INC.
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
Zhao QA, Zhou P, Goodrich R, Drmanac RT;
WPI: 2001-442253/47.
P-PSDB: AAM40125.
Novel nucleic acids and polypeptides, useful for treating disorders
such as central nervous system injuries -
Claim 1: SEQ ID NO 1484; 10078pp; English.
The invention relates to human nucleic acids (AA157798-AA161369) and
the encoded polypeptides (AAM38642-AAM42213) with nootropic,
immunosuppressant and cytostatic activity. The polynucleotides are useful
in gene therapy. A composition containing a polypeptide or polynucleotide
of the invention may be used to treat diseases of the peripheral nervous
system, such as peripheral nervous injuries, peripheral neuropathy and
localised neuropathies and central nervous system diseases, such as
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
utilisation of the activities such as: Immune system suppression,
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
and thrombolytic activity, cancer diagnosis and therapy, drug screening,
assays for receptor activity, arthritis and inflammation, leukaemias and
C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
specification.
Sequence 3622 BP; 1143 A; 656 C; 657 G; 1166 T; 0 other;

Db 431 ttgtcatctttgacggtttgagggaaacttttgggagccctatcatacctgcacttcaagc 490
Qy 361 aagtattttaaacttcaactactgagaccacacaaatggtagaagccttactgtttgggca 420
Db 491 aagtattttaaacttcaactactgagaccacacaaatggtagaagccttactgtttgggca 550
Qy 421 tctactcatatgtcacagcttcttgacattactaaaattgtgtgatgttcagccaatgcag 480
Db 551 tctactcatatgtcacagcttcttgacattactaaaattgtgtgatgttcagccaatgcag 610
Qy 481 tatttgacctgacttgcagctctctgggcaagcagaagtggagcgagcatcttctt 540
Db 611 tatttgacctgacttgcagctctctgggcaagcagaagtggagcgagcatcttctt 670
Qy 541 cttaaggtatggatggcaccaggacaccatttccattcttgagagtgcttaatacaagac 600
Db 671 cttaaggtatggatggcaccaggacaccatttccattcttgagagtgcttaatacaagac 730
Qy 601 ctgttcttgaagtgatttaagtccacatccatcggtcacaaaatctgcacaatagacatt 660
Db 731 ctgttcttgaagtgatttaagtccacatccatcggtcacaaaatctgcacaatagacatt 790
Qy 661 ttagtctacgataaccatcttcatgtggaagatctctgaaggttggaggtttcttctaga 720
Db 791 ttagtctacgataaccatcttcatgtggaagatctctgaaggttggaggtttcttctaga 850
Qy 721 atctatagccttcataccacacttcaatcaatgaattcagagaaatcagacaatgttaagt 780
Db 851 atctatagccttcataccacacttcaatcaatgaattcagagaaatcagacaatgttaagt 910
Qy 781 tttagagtttcatcttcatggaggtaccaggttacggttcggtggaatcaggggttttgcagaa 840
Db 911 tttagagtttcatcttcatggaggtaccaggttacggttcggtggaatcaggggttttgcagaa 970
Qy 841 agtaactctgattgtggtacactgcaactgaaaagatttgaatctgcaaaatttgacagccaat 900
Db 971 agtaactctgattgtggtacactgcaactgaaaagatttgaatctgcaaaatttgacagccaat 1030
Qy 901 cagcattcagatgttatctgttcaatcagaacctgacgacagctttccaagcttggatca 960
Db 1031 cagcattcagatgttatctgttcaatcagaacctgacgacagctttccaagcttggatca 1090
Qy 961 gtatcatatcagagtgagaagatgtcaacagctatctgtcacaaatcttaccagatcat 1020
Db 1091 gtatcatatcagagtgagaagatgtcaacagctatctgtcacaaatcttaccagatcat 1150
Qy 1021 cagtatattggagagagaccactatgtgccattttgaaacaaaagcttcccaacaatc 1080
Db 1151 cagtatattggagagagaccactatgtgccattttgaaacaaaagcttcccaacaatc 1210
Qy 1081 cgcattccagagcaaaattgaggttcataaagcccgagaagactatttcagctgtttaaactt 1140
Db 1211 cgcattccagagcaaaattgaggttcataaagcccgagaagactatttcagctgtttaaactt 1270
Qy 1141 cattgccctaaatgtcatttggctgcaagaagtccacatgagggcgatttgataaatt 1200
Db 1271 cattgccctaaatgtcatttggctgcaagaagtccacatgagggcgatttgataaatt 1330
Qy 1201 ttctcaggtggtgcaactaaaaccccgattgtcaagttacaaaatcacatcattatgat 1260
Db 1331 ttctcaggtggtgcaactaaaaccccgattgtcaagttacaaaatcacatcattatgat 1390
Qy 1261 tcacaaaatctgacacactaaaatcaaaaagcagaaaagtagcaggttcatatttggaaa 1320
Db 1391 tcacaaaatctgacacactaaaatcaaaaagcagaaaagtagcaggttcatatttggaaa 1450
Qy 1321 aataatggttatctcccgctttcaaatgaatgtctacttttgatagaagagggtacactc 1380
Db 1451 aataatggttatctcccgctttcaaatgaatgtctacttttgatagaagagggtacactc 1510
Qy 1381 agtgaatttgcacaaactctgcacaaaggtttaaattagtgtaatttctgtgagatctggccac 1440
Db 1511 agtgaatttgcacaaactctgcacaaaggtttaaattagtgtaatttctgtgagatctggccac 1570

|||||
Db 971 agtaactctgctggaactcaactgaaaggattgagatctgcaaatgtgacagccaat 1030
QY 901 cagcattcagatgttatctgtcaatcagaacctgacagctttccaagctctggatca 960
|||||
Db 1031 cagcattcagatgttatctgtcaatcagaacctgacagctttccaagctctggatca 1090
QY 961 gtatcatatcacaggtgagaagatgtcaacagctctctgtacataacttacagatcat 1020
|||||
Db 1091 gtatcatatcacaggtgagaagatgtcaacagctctctgtacataacttacagatcat 1150
QY 1021 cagtatgttgagagacacacctatgtgccattttgaaacaaaaagctctccaacaatc 1080
|||||
Db 1151 cagtatgttgagagacacacctatgtgccattttgaaacaaaaagctctccaacaatc 1210
QY 1081 cgcattcagagcaaaattgaggtcatataagccacagaagactatttcagctctgtaaaatt 1140
|||||
Db 1211 cgcattcagagcaaaattgaggtcatataagccacagaagactatttcagctctgtaaaatt 1270
QY 1141 cattgccctaaatgtcatttctgctcaagaagtccacatgagggcgatcttgatataatt 1200
|||||
Db 1271 cattgccctaaatgtcatttctgctcaagaagtccacatgagggcgatcttgatataatt 1330
QY 1201 ttctcaggtggtgcactaaaaacccagttgtcaagttacaaaatcacatcattatgat 1260
|||||
Db 1331 ttctcaggtggtgcactaaaaacccaggtgtcaagttacaaaatcacatcattatgat 1390
QY 1261 tcaaaaaatgcagaccactaaaaatcaaaaaagcagaagaagtagcaggttcattttgtgaaa 1320
|||||
Db 1391 tcaaaaaatgcagaccactaaaaatcaaaaaagcagaagaagtagcaggttcattttgtgaaa 1450
QY 1321 aataatggtattctccgctttccaatgaatgtctacttttgatagaaggaggtacactc 1380
|||||
Db 1451 aataatggtattctccgctttccaatgaatgtctacttttgatagaaggaggtacactc 1510
QY 1381 agtgaatttgcaactctgcagaactgaagaagtttaatagtgtaattccctgtgagatctggccac 1440
|||||
Db 1511 agtgaatttgcaactctgcagaactgaagaagtttaatagtgtaattccctgtgagatctggccac 1570
QY 1441 gaagacctggaaactttgacacttccagcaccattctctatacagaaggacaatacatcac 1500
|||||
Db 1571 gaagacctggaaactttgacacttccagcaccattctctatacagaaggacaatacatcac 1630
QY 1501 tatgatgtaaacagtgcttagtttgatgagatccatacaaaaatctaaatccctgtgattgat 1560
|||||
Db 1631 tatgatgtaaacagtgcttagtttgatgagatccatacaaaaatctaaatccctgtgattgat 1690
QY 1561 aaaaatcgtgattcccttctctgtgcaagaagcactgggtattgtaccctccaatat 1620
|||||
Db 1691 aaaaatcgtgattcccttctctgtgcaagaagcactgggtattgtaccctccaatat 1750
QY 1621 gtgtttgttatgacctttacacttgatgatggaacagagtagtactagaagcctatctcatg 1680
|||||
Db 1751 gtgtttgttatgacctttacacttgatgatggaacagagtagtactagaagcctatctcatg 1810
QY 1681 gatttgcacaaattctccagattccagcatcagaagttctgattgagatgagaccttcag 1740
|||||
Db 1811 gatttgcacaaattctccagattccagcatcagaagttctgattgagatgagaccttcag 1870
QY 1741 aaaaagtggtgatgatgatgatgtgtttgtctccaggaataaaaattgagtcacat 1800
|||||
Db 1871 aaaaagtggtgatgatgatgatgtgtttgtctccaggaataaaaattgagtcacat 1930
QY 1801 ccggtggttggaaatgtctcatcaatgatacaatgtcacaaatgggaacagagatacaaat 1860
|||||
Db 1931 ccggtggttggaaatgtctcatcaatgatacaatgtcacaaatgggaacagagatacaaat 1990
QY 1861 tgcattcagattttgacaccacagtttcagagaagatgtaattctaa 1905
|||||
Db 1991 tgcattcagattttgacaccacagtttcagagaagatgtaattctaa 2035

AAH17603
ID AAH17603 standard; cDNA; 2383 BP.
XX
AC AAH17603;
XX
DT 26-JUN-2001 (first entry)
XX
XX Human cDNA sequence SEQ ID NO:17112.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX
XX full-length cDNAs defined in the specification, and for the detection
XX
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 17112; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
XX
XX full-length cDNAs defined in the specification. Where a primer set
XX
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX
XX to the complementary strand of a polynucleotide which comprises one of
XX
XX the 5602 nucleotide sequences defined in the specification, where the
XX
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX
XX of an oligonucleotide comprising a sequence complementary to the
XX
XX complementary strand of a polynucleotide which comprises a 5'-end
XX
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX
XX polynucleotide which comprises a 3'-end sequence, where the
XX
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX
XX the specification. The primer sets can be used in antisense therapy and
XX
XX in gene therapy. The primers are useful for synthesising polynucleotides,
XX
XX particularly full-length cDNAs. The primers are also useful for the
XX
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX
XX AAH13633 to AAH18742 represent human cDNA sequences; AA892446 to
XX
XX AA895893 represent human amino acid sequences; and AAH13629 to AAH13632
XX
XX represent oligonucleotides, all of which are used in the exemplification
XX
XX of the present invention.
XX
XX Sequence 2383 BP; 750 A; 456 C; 459 G; 718 T; 0 other;

Query Match 99.8%; Score 1901.8; DB 22; Length 2383;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1903; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 atgtcttgggtccagcaacaaattatatatatatacccccgaatcaacttaagggtggt 60
|||||
Db 439 atgtcttgggtccagcaacaaattatatatatatacccccgaatcaacttaagggtggt 498

PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 XX full-length cDNAs -
 PS
 XX Claim 8: SEQ ID 11187; 2537pp + CD ROM; English.
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 2631 BP; 826 A; 473 C; 491 G; 841 T; 0 other;

Query Match 100.0%; Score 1905; DB 22; Length 2631;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1905; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 atgtcttgggtccagcaacaattatatatatatacacccctgaatcaacttaagggtggt 60
 Db 24 atgtcttgggtccagcaacaattatatatatatacacccctgaatcaacttaagggtggt 83
 Qy 61 acaattgcaatgctatggtgtgtgaagttctttaagcccccatatcttaagcaaaagga 120
 Db 84 acaattgcaatgctatggtgtgtgaagttctttaagcccccatatcttaagcaaaagga 143
 Qy 121 actgattattgctcagttgtaactattgtggaccagacaaatgtataaactaacttgctcg 180
 Db 144 actgattattgctcagttgtaactattgtggaccagacaaatgtataaactaacttgctcg 203
 Qy 181 ctcttagtggaaactatgaagccctcccaataatttataaaatggagatatgttctgcg 240
 Db 204 ctcttagtggaaactatgaagccctcccaataatttataaaatggagatatgttctgcg 263
 Qy 241 ttccacagctgaagattcaagtatataaaaggagactcaggttatcaccagctctggc 300
 Db 264 ttccacagctgaagattcaagtatataaaaggagactcaggttatcaccagctctggc 323
 Qy 301 ttfgcatctttgacgttttgagggaactttggagccctcctatcacactcgcaacttcaagc 360
 Db 324 ttfgcatctttgacgttttgagggaactttggagccctcctatcacactcgcaacttcaagc 383
 Qy 361 aagattttaacttcaactactgaggaacacaaaatggtagaagccttaagtttgggca 420
 Db 384 aagattttaacttcaactactgaggaacacaaaatggtagaagccttaagtttgggca 443
 Qy 421 tctactcatatgcacgtcttgacattactataaaatggtgtagtttcagcccaatgcag 480
 Db 444 tctactcatatgcacgtcttgacattactataaaatggtgtagtttcagcccaatgcag 503
 Qy 481 tatttgacctgacttgcagctcttgggcaaaagcagaagtggaagcagcatcttctt 540
 Db 504 tatttgacctgacttgcagctcttgggcaaaagcagaagtggaagcagcatcttctt 563
 Qy 541 ctaaagggtatggatggcaccaggaacacatttccatcttggagaggtcttaatacaagac 600

Db 564 ctaaagggtatggatggcaccaggaacacatttccatctctgggaggtcttaatacaagac 623
 Qy 601 cttgtcttgaaggtgatttaagtccatccatcggtcacaaaatctgacaatagacatt 660
 Db 624 cttgtcttgaaggtgatttaagtccatccatcggtcacaaaatctgacaatagacatt 683
 Qy 661 ttaagtccagataaccatgttctatgtgcaagatctctgaagtttgaaagcttctttaga 720
 Db 684 ttaagtccagataaccatgttctatgtgcaagatctctgaagtttgaaagcttctttaga 743
 Qy 721 atctatagcttccataccaaaacttcaatcaatgaattcagagaaatcagacaattgtaagt 780
 Db 744 atctatagcttccataccaaaacttcaatcaatgaattcagagaaatcagacaattgtaagt 803
 Qy 781 ttaagtttcatctctatggaggtaccagttacagttacggttcggggaatcaggtcttgcagaa 840
 Db 804 ttaagtttcatctctatggaggtaccagttacggttcggggaatcaggtcttgcagaa 863
 Qy 841 agtaactctgattggtacaaactgaaaagatttagaactctcaaatcttgacagcacaat 900
 Db 864 agtaactctgattggtacaaactgaaaagatttagaactctcaaatcttgacagcacaat 923
 Qy 901 cagcattcagatgttatttgcatacagaacctgacgacagcttccaagctctggatca 960
 Db 924 cagcattcagatgttatttgcatacagaacctgacgacagcttccaagctctggatca 983
 Qy 961 gtaaatattacagaggtagaagaatgtcaacagctatctgctacaatacttacaagatcat 1020
 Db 984 gtaaatattacagaggtagaagaatgtcaacagctatctgctacaatacttacaagatcat 1043
 Qy 1021 cagtatttggagaggaacaccatctatgtccattttgaaacaaaagctctccaacaatcac 1080
 Db 1044 cagtatttggagaggaacaccatctatgtccattttgaaacaaaagctctccaacaatcac 1103
 Qy 1081 cgcctcagcagcaaatggaggtcatabataagcccagagaactatttcagctctgttaaaactt 1140
 Db 1104 cgcctcagcagcaaatggaggtcatabataagcccagagaactatttcagctctgttaaaactt 1163
 Qy 1141 cattgcctaaatgctcatttctgctcaagaagtctccacatgagggcgtttggatataaatt 1200
 Db 1164 cattgcctaaatgctcatttctgctcaagaagtctccacatgagggcgtttggatataaatt 1223
 Qy 1201 ttccagatgggtcgaactaaaaccccgcttccaagtgtcacttttgatagaaggaggtacactc 1260
 Db 1224 ttccagatgggtcgaactaaaaccccgcttccaagtgtcacttttgatagaaggaggtacactc 1283
 Qy 1261 tcaaaaatctggagacactaaaatcaaaaaggacgaaaagtagcagttcatttttggtaaa 1320
 Db 1284 tcaaaaatctggagacactaaaatcaaaaaggacgaaaagtagcagttcatttttggtaaa 1343
 Qy 1321 aataatggtattctcccgcttccaagtgtcacttttgatagaaggaggtacactc 1380
 Db 1344 aataatggtattctcccgcttccaagtgtcacttttgatagaaggaggtacactc 1403
 Qy 1381 agtgaatttgcacaaactctgaacaaagttaataagttaattctctgagatcttggccac 1440
 Db 1404 agtgaatttgcacaaactctgaacaaagttaataagttaattctctgagatcttggccac 1463
 Qy 1441 gaagacctgggaacttttgacacttccagaccattctctatacaaggaaacaatacatcac 1500
 Db 1464 gaagacctgggaacttttgacacttccagaccattctctatacaaggaaacaatacatcac 1523
 Qy 1501 tatggatgaacaggtgttcttagtttgagatccatacaaaaatctaaatctccctggttga 1560
 Db 1524 tatggatgaacaggtgttcttagtttgagatccatacaaaaatctaaatctccctggttga 1583
 Qy 1561 aaaaactcgtgattctctctctctggtgcaagaacactgggtattgtaccctccaatcat 1620
 Db 1584 aaaaactcgtgattctctctctctggtgcaagaacactgggtattgtaccctccaatcat 1643
 Qy 1621 gtgtgttattgacactttacacttgatgatgaacagagtagtactagaagccttatctcatg 1680

PT peripheral neuropathy -

XX

PS Example 2; Page 90-93; 126pp; English.

XX

CC The present sequence represents module 8 of the FK-520 polyketide
CC synthase (PKS) gene cluster of strain MA6548. FK-506 is a potent
CC immunosuppressant, and acts through initial formation of an
CC intermediate complex with protein immunophilins known as FK-506
CC binding proteins. The nucleic acids are used for producing polyketide
CC compounds. The polyketide compounds can be used as immunosuppressants to
CC prevent or treat transplant rejection, graft-versus-host disease or
CC uveitis. They can also be used for treating e.g. alopecia universalis,
CC autoimmune chronic active hepatitis, inflammatory bowel disease,
CC multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
CC also have neurotrophic activity and can be used to promote neurite
CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC and in intact animals, they promote regrowth of damaged facial and
CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC the brain. They can also be used for treating e.g. Parkinson's disease,
CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.

XX

SQ Sequence 1574 AA;

Query Match 3.2%; Score 105.5; DB 21; Length 1574;
Best Local Similarity 21.9%; Pred. No. 2.2;
Matches 92; Conservative 44; Mismatches 160; Indels 125; Gaps 21;

QY 167 OLLGKAEVDGA-----SFLKVDGTRTPPSHRVLI----QDLVLEGDLSHIH 211
DB 1130 qvfveevdsdgfvahpdlldavfsavgdgsrqp-tgwrldavhasdatviracit--- 1185
QY 212 RLQNLTDILVDN---HVHVARSLKVGSLRIYSLSLHTKLQSMNSNQTMLSLEFHLGG 268
DB 1186 rrdsgvvlaafagagmpvltaesvtlg-----evasagsgdesdglrllew----- 1232
QY 269 TSYGRGIRVLPENSVDQLKLDLESANLTANOHSVDVICQSEPDSDFFSPSGSVSLYEVE 328
DB 1233 -----lpvaeahydgadelpegytlitathpd-----dpddptnphmtpttrhtqt 1278
QY 329 COOLSA-----TILTDHOYLERT----PLCAI--LKOKAPQOYRIRAKL----RSYKPRRL 374
DB 1279 trvltalqhlitntltvhttdppgaavtgltrtaqnehpgrihliethhphptplpl 1338
QY 375 FQSVKLHCPKCHLLQEPHGGDLIIIFQDGATKTPVVKLQNTSLYDSKIWTNKNQGRKV 434
DB 1339 tqittlhqphlrlnntltlthpl-----tpitthhntt-----ttptntp---- 1378
QY 435 AVHFVKNNGTLPNSCELLIEGGLTSEICKLSNKFNSVIPVRSGHEDLELLELDS-APFL 493
DB 1379 -----plnphallitgg-----sgtlaqilarhlnhphtyllsrtppppt 1419
QY 494 IOGTIHHYCKQCSLSRLNSLNSLDKTSWIPSSVAEALGIV--PLQYVFWMTFLDDG 551
DB 1420 tpqt--hpc-----dltd-----ptqitqalthipqplgtifgtaatldda 1459
QY 552 T 552
DB 1460 t 1460

Search completed: April 9, 2002, 17:08:10

Job time: 589 sec

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160983.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.2%; Score 106; DB 21; Length 1123;

Best Local Similarity 20.4%; Pred. No. 1.1; 169; Indels 172; Gaps 25;

Matches 104; Conservative 65; Mismatches 65; Mismatches 172; Gaps 25;

Oy 118 TSSKYFNFTEDHKVMEAL---RVMASTHMSPSWTLLKLCVQPMQYFDLT----- 165

Db 62 tpcnwfgitcdsknvasinftrsvsglqp-----eigelsqildstnnsgtip 116

Oy 166 -----COLLGKAEVGGASFLKXWGDTRTPFSSWRVLQDL-VLEGDL-----HHRLO- 214

Db 117 stlgnctkiatlidsengfskipd-tidslkrlevlylnflgtelpeslfrpkiqv 175

Oy 215 -----NLT-----IDILVDNHV--HVARSKVGSLFIYSLSH----- 245

Db 176 lyldynnltpipqsigdakelvelsmyanqfsgnipesignssslqilyhrnkivgsl 235

Oy 246 -----TKLQSMNSNQTMLSLEPH-LHGQTSYGRGIRVLPES 281

Db 236 peslnllgnltlrfvgnnslqgvrfgspcnklltldlsynefeggyppalg-----n 289

Oy 282 NSDVDQL-----KKDLESANLTANOHSVDVICQSEPDSPFSGSVSLY 324

Db 290 cssldalvivsgnlsgtspsslgmknltlhlsefr-----lsgsipa- 333

Oy 325 EVERCOOLSATILTDHOYLERTPLCAILKQKAPQOYRIKALRSYKPRRLFOSVKLHCPK 384

Db 334 elgncssnllklnndqlvggipsal-----gkrlkleslfe----- 373

Oy 385 CHLLQEVPHGDLDFODGATKTPVKLONTSLYDSKIWTTKNOKRKVAVHVKNN-- 442

Db 374 -----rfsgeipielwksqitqllyvgnltgelpvemtemkk-ikiatlf--nnsf 423

Oy 443 -GILPLSNECLLLIEGGLTSEICKLSKNFNSVIPVRSGH-EDLELDSLAPFL---IOGT 497

Db 424 ygaippg-----lgnssleevdfignkltgeipponlchgrklrlnlgsnllhgtipas 478

Oy 498 IHHYGCKQCSSLSRI-----QNLNSLVDKTS 523

Db 479 igh-----Cktirrfilrennislpefs 503

RESULT 14

[illegible]

Qy	291	DLESANLTANOHSDVICOSEPDDSPSSGVSLSYEVERCOOLSATI--LTDHOYL-----	343
Db	1197	nisscrvtven---dvmkryeklgifp-----hikvidtlsaefpaltynilytyqgg	1245
Qy	344	ERTPLCAILKOKAPOQYRIRAKLRSYKPRRLFQSVKLHCPCKHLAQEPHGGDLIIFQD	403
Db	1246	ehdvlpmlmkrk-----kictlnnkrnankkkvhv-knhlynevvdakdtqihken	1295
Qy	404	GATKTPVVKLQNTSLYDSKIIWTTNQGRKRVAVHVFVKNNGILPLSNCELLLIIEGGTLSEI	463
Db	1296	nnnnn-----mmsgnvencklnkesygnssncintnlnlennic-----hdisinkn	1346
Qy	464	CK--LSNKNFSVIPVRSGHEDLELDLSDAPFLIOGTIIHYGCKQCS-----SLRSIQNL	515
Db	1347	ikvtlnnsmnsi-----snnenve-tnlncvseragshhiygkeeksigsadtnilsaqns	1401
Qy	516	NS 517	
Db	1402	nn 1403	
RESULT 12			
AAB18793			
ID	AAB18793 standard; Protein: 1139 AA.		
XX	AAB18793:		
XX	22-JAN-2001 (first entry)		
XX	Amino acid sequence of a verticillium (Vel.1) polypeptide.		
XX	Verticillium polypeptide; Vel.1; Vel.2; resistance gene;		
KW	fungal resistance; verticillium wilt; vascular disease; transgenic plant.		
XX	Lycopersicon esculentum.		
XX	WO200055336-A2.		
XX	21-SEP-2000.		
XX	24-FEB-2000; 2000WO-CA00184.		
XX	12-MAR-1999; 99US-0124129.		
PR	22-APR-1999; 99US-0130586.		
XX	(MIAC) CANADA DEPT AGRIC & AGRI-FOOD CANADA.		
XX	Kawchuk LM, Lynch DR, Hachey J, Kulcsar F;		
XX	WPI: 2000-618919/59.		
DR	N-PSDB: AAA75858.		
XX	New Verticillium polynucleotide useful for producing a transgenic plant		
PT	such as a tomato, potato, hop, alfalfa, cucumber, groundnut, horse		
PT	radish, tobacco, or red pepper plant, that is resistant to Verticillium		
PT	species		
XX	Claim 8; Page 66-69; 102pp; English.		
PS			
XX	The present sequence represents a tomato verticillium polypeptide. The		
CC	polypeptide is encoded by vel.1 and Vel.2 resistance genes. The		
CC	polypeptide confers resistance to verticillium species in a plant.		
CC	Verticillium are fungi, which cause verticillium wilt in plants. This		
CC	is a common vascular diseases that causes sever yield and quantity		
CC	losses in many crops. The polypeptide is useful for producing a		
CC	transgenic plants that are resistant to Verticillium species.		
XX	Sequence 1139 AA:		
XX	SQ		

Query Match 3.3%; Score 111; DB 21; Length 1139;
Best Local Similarity 20.7%;
Pred. No. 0.37;

```

Db 193 tsavf---vwdgtadp-pa----- 207
Oy 235 VGSFLRIYSLTKKQSMNSENOTMLSEFLHGGTSYGRGIRVLPESNSVDVQLKKDLES 294
Db 208 -----silakseedkafslsvh---tllsr----- 231
Oy 295 ANLTANQHSVDVTCQSEPDSPSSGVSLSYEVERCQQLSATILTDHQLYLERIPICAILKQ 354
Db 232 -----dvll-----sftvgtllrvh-----lshlf----- 253
Oy 355 KAPOQYRIRAKLRSYKPRRLFSVKLHCPKCHLLQVPEHGGDLIIFQDGAIKTPVVKLQ 414
Db 254 -----yrakpgdw-----vklv---hllec----- 271
Oy 415 NTSLYDSKIMTTKQKGRVAVHVKNGILPLSNECLLIEGGTLSEICKLSKNFNSVI 474
Db 272 -----drsgsvikvtsstkvh-hlaqddrlv-----ekimrydkrlss--klghisfwcf 319
Oy 475 PVRSCHDELLELDLAPFLIOGTIIHHYCKQCSSLRSIONLSLYDKTSWIPSSVAEALG 534
Db 320 pspggitetd--dnacpvrilmlditfpkvtc-----kyrcivrvvaaypwqvedfcs 370
Oy 535 IYPLQYFVMTFTLDGDTGVLEAYL--MDSKRFQIPASEVLMDLQKSVDMIMDMFCP 592
Db 371 denrrhhqvl-ltledstatleafcnkdaeyfwlgfqdt---etlkrknwll----- 421
Oy 593 PGKIKDAY-----PWLECFIKSNVNTGTD--NOICVQIFDT 627
Db 422 -giressnfvapnppwleicilsy-ytnkadpwntrlyrifgt 463

RESULT 9
AAR88122
ID AAR88122 standard; Protein; 1144 AA.
AC
XX
XX
DT 28-MAR-1996 (first entry)
XX
DE Tobacco mosaic virus resistance N gene protein.
XX
KW Tobacco mosaic virus resistance; TMV; N gene; Solanaceae:
KR crop improvement; transgenic plant; crop improvement.
XX
OS Nicotiana glutinosa.
XX
FH Key Location/Qualifiers
FT 1..150
FT /label= Cytoplasmic_region
FT Binding-site 216..224
FT /label= P-loop
FT /note= "ATP/GTP-binding site motif"
FT Binding-site 228..229
FT /label= P-loop
FT /note= "ATP/GTP binding site motif"
FT Binding-site 297..302
FT /label= P-loop
FT /note= "ATP/GTP binding site motif"
FT Region 590..928
FT /label= Leucine-rich_region
FT /note= "the leucine-rich region (aa 590-928)
FT includes 13 repeats of approx. 25 aa
FT length"
XX
PN W0535024-A1.
XX
PD 28-DEC-1995.
XX
PF 16-JUN-1995; 95WO-US07754.
XX
PR 17-JUN-1994; 94US-0263663.
XX

```

```

PA (REGC ) UNIV CALIFORNIA.
PA (USDA ) US SEC OF AGRIC.
XX Baker BJ, Whitham SA;
XX WPI: 1996-058144/06.
DR N-PSDB; AAT09340.
XX Plant virus resistance gene N sequences from tobacco - useful for
PT generating transgenic Solanaceous plants resistant to Tobacco Mosaic
PT Virus
XX Example 6; Page 52-60; 98pp; English.
XX The Nicotiana glutinosa N gene protein (AAR88123) mediates
CC resistance to tobacco mosaic virus (TMV). The gene (AAT09341)
CC coding for the protein was obt'd. from a N. glutinosa leaf genomic
CC library by screening with a cDNA clone. DNA sequences encoding the
CC protein can be used to generate transgenic plants, esp. Solanaceae,
CC resistant to TMV.
XX Sequence 1144 AA;
SQ
Query Match 3.7%; Score 122; DB 17; Length 1144;
Best Local Similarity 18.2%; Pred. No. 0.03;
Matches 135; Conservative 106; Mismatches 200; Indels 300; Gaps 36;
Oy 71 IYKNGDIVRPHRL-----KIQVKKETOG--ITSSGPFASLTPEFTLGAPITPRTSSKTFN 124
Db 335 liekndliyevalpdhesiqfkhafgkevpnenfeklslevnyakglpl----- 387
Oy 125 FTTEPHKMWVEALRWAS-----THM---SPSWTLKKL-----CDVQPMQ--- 160
Db 388 -----alkvgsgllhnlrltewksaiehmknnsygiidkkisydglepkqgem 437
Oy 161 YFDLTCLLQKAEVDGASFLKAWD-----GTRTPFPWSRVLIQDLVLEGDSLHHR 212
Db 438 fldiaacflrgeek---dyilqileschigaeyglr-----illdkslvfise 481
Oy 213 LONLTIDILVVDNHVVARSLK-VGSFLRIYSLTKKQSMNSENOTMLSEFLHGGTSY 271
Db 482 ynqvqmhdlidgmkyivnfkdpgrsrilw-lakeveevmsnntgtname--alwvsy 538
Oy 272 GRGIRVLPESNDVDQLKKDLESANL-TANQHSDV-----ICQSEPDSPSSGGS 320
Db 539 sstlrf---snqavknmkr-lrvfnmgrssthaidyipnlnrcfvctnypwesfstfe 594
Oy 321 -----VSLYEVR 328
Db 595 lkmlvhlglrhnsrlhwtetkhlpslrridslwskrltrtpdftgmpnleynlyqcsn 654
Oy 329 -----COOLSATILTDHQLYLERIPICAILK-----OKAPOQY-RIRA 364
Db 655 leevhhsigccskviglylndckslkrfp-cvnvesleylgrscdslekipeiygrmkp 713
Oy 365 KLRYSK-----PRRLFSQSVKLHCPK-----CHL 387
Db 714 eiqlhmqqsgirelpssifq-ykthvtklllwnmknvalpssicrlkslvalsvsgcsk 772
Oy 388 LOEVPEH-GDLDIIFQDGAIKTPVVKLQNTSLYDSKIRTKQKGRVAVHVKNGILP 446
Db 773 lesipeeigdlunirvfdasdtllirppssiirlhkl-illmfrgfkdvhf----- 823
Oy 447 LSNECLLLIEGCTLSEICKLSKNFNSVIPVRSCHDELLELDLAPFLIOGTIIHHYCKQC 506
Db 824 -----efppvaeghlshleylnisyncnliidggip-----esi 854
Oy 507 SSLRSIONLSLVDKTSWIPSSVAEALGIVPLOYVFMFTFTLDGDTGVLEAY-LMDSKRF 565
Db 855 gslssikkidlsrnnfehlpssiaq-----lgaigsidlkacqrl 894
Oy 566 FOIPASEVLMDLQKSVDMIMDMFCPGIK---IDAYPWLECFIKSYNWT----- 613

```

Query Match 10.8%; Score 358; DB 22; Length 66;
Best Local Similarity 100.0%; Pred. No. 1.1e-27;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 569 PASEVLMDDDLQKSYDMIMDFCPGKIKIDAYPWLECFIKSYNTGTDNQCICQIFDIT 628
|||||
Db 1 pasevlmdddlqskysdmiimdfcpvgkikidaypwlecfiksyntgtdnqicqifdt 60

OY 629 VAEDVI 634
|||||
Db 61 vaedvi 66

RESULT 7
AAM34264
ID AAM34264 standard; Protein: 30 AA.
XX AC AAM34264;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #8301 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX OS genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207450.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WI WI; 2001-488897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta -
XX PS Claim 27; SEQ ID No 34533; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENP;
XX CC see A131315-AA157546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders.
XX SQ Sequence 30 AA;

Query Match 4.4%; Score 147; DB 22; Length 30;
Best Local Similarity 96.7%; Pred. No. 1.9e-07;
Matches 29; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 389 QEVPHGGDLIIIFODCATKTPVVKLQNTSL 418
|||||
Db 1 qevphgdlilifqgcatktpdvklqntsl 30

RESULT 8

AAG58505
ID AAC58505 standard; Protein: 467 AA.
XX AC AAG58505;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 75533.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 23-MAR-1999; 99US-0123548.
XX PR 29-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 04-MAY-1999; 99US-0132407.
XX PR 05-MAY-1999; 99US-0132484.
XX PR 06-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 07-MAY-1999; 99US-0132487.
XX PR 11-MAY-1999; 99US-0132863.
XX PR 14-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 18-MAY-1999; 99US-0134370.
XX PR 19-MAY-1999; 99US-0134768.
XX PR 20-MAY-1999; 99US-0134941.
XX PR 21-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.

```
SQ Sequence 634 AA:
Query Match 99.8%; Score 3317; DB 22; Length 634;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 633; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSLVPATNYIYTPNLKGGTIVNVYGVVKKFPPYLSKGTIDYCSVVTIVDQTNVLTCL 60
DB 1 mslvpatnyiytpnlqkggtivnvvgvkvffkppylskgtdycsvvtivdgtvnkltcl 60
QY 61 LFSGNYEALPIIYKNGDIYRFHRLKIQYKKETQGTSSGFSASLTGEGTLGAPIIPRTSS 120
DB 61 lfsngyealpilykngdivrfhrklqvykketqgtssgsfaaltfegtlgapiiprtss 120
QY 121 KYNFETDEHKMVEALRVWASTHMSPTLLKLCVOPMQYFDLTCLLGKAEVDCASFL 180
DB 121 kynfetdehkmealrvwasthmspswtllkldcvpmqyfdltcqlgkaevdgasfl 180
QY 181 LKWDGTRTPFPSSRWLVLDVLEGLSHIHLQNLITDILVYDNHVVARSILKVGSLR 240
DB 181 lkwdgtrtpfswrvlqdlvlegdshihrlqnlitdilvydnhhvvarsilkvgsflr 240
QY 241 IYSLHTKLOSMNSNOTMLSEFLHGGTSYGRGIRVLPESNSDVQLKKDLESANLTAN 300
DB 241 iyslhtklqsmnsenqtmlslefhlhggtsygrgirvlpesnsdvdqkkdlesanltan 300
QY 301 QHSDVICQSEPDSPSSGVSLSYEVERCOOLSATILTDHOYLERTPLCAILKQKAPQY 360
DB 301 qhsdvlcseqpddsfpsgsvslyevercqolsatiltldhoylerltpcailkqkappqy 360
QY 361 RIRAKLSYPRRLFOSVKLHCPKCHLLQEPVHPHGGDIIFQDGATKTPVVKLQNTSLYD 420
DB 361 riraklsyprrlfsvklhpcpkchllqevhphegdliifdgatktpdvklqntslyd 420
QY 421 SKIWTNKGKRVAVHFVNNGILPLSNCELLIEGGTISEICKLSKNFNSVIPVRSKH 480
DB 421 sklwtknkgkrvavhfvnngilplsnecellieggtlseicklsknfnsvipvrsgh 480
QY 481 EDLELIDLSAPFLIOGTIHHYCKQCSSLRSLQNLNSLDVKTSMIPSSVAEALGIPLQY 540
DB 481 edlelidlsapfliogtthhyckqcsslrslnslqnlslvdktswipssvaealgiplqy 540
QY 541 VFVMTITLDGTVLEAYLMDSDKFFQIPASEVLMDDDLQKSVDMIMDFCPPGIKIDAY 600
DB 541 vfmvtitldgvtvleaylmdsdckffqipasevlmdddlqksvdmimdmfcpvgikiday 600
QY 601 PWLECFIKSVNTNGTDNOICYOIFDITVAEDVI 634
DB 601 pwlecfiksvntngtdnqicyoifdittvaedvi 634

RESULT 3
AAB93478
ID AAB93478 standard; Protein; 634 AA.
XX
AC AAB93478;
XX
XX 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12761.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
XX
PD 07-FEB-2001.
XX
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
XX
PR 29-JUL-1999; 99JP-0248036.
```

```
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs.
PS Claim 8: SEQ ID 12761; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX Sequence 634 AA:
SQ
Query Match 99.8%; Score 3317; DB 22; Length 634;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 633; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSLVPATNYIYTPNLKGGTIVNVYGVVKKFPPYLSKGTIDYCSVVTIVDQTNVLTCL 60
DB 1 mslvpatnyiytpnlqkggtivnvvgvkvffkppylskgtdycsvvtivdgtvnkltcl 60
QY 61 LFSGNYEALPIIYKNGDIYRFHRLKIQYKKETQGTSSGFSASLTGEGTLGAPIIPRTSS 120
DB 61 lfsngyealpilykngdivrfhrklqvykketqgtssgsfaaltfegtlgapiiprtss 120
QY 121 KYNFETDEHKMVEALRVWASTHMSPTLLKLCVOPMQYFDLTCLLGKAEVDCASFL 180
DB 121 kynfetdehkmealrvwasthmspswtllkldcvpmqyfdltcqlgkaevdgasfl 180
QY 181 LKWDGTRTPFPSSRWLVLDVLEGLSHIHLQNLITDILVYDNHVVARSILKVGSLR 240
DB 181 lkwdgtrtpfswrvlqdlvlegdshihrlqnlitdilvydnhhvvarsilkvgsflr 240
QY 241 IYSLHTKLOSMNSNOTMLSEFLHGGTSYGRGIRVLPESNSDVQLKKDLESANLTAN 300
DB 241 iyslhtklqsmnsenqtmlslefhlhggtsygrgirvlpesnsdvdqkkdlesanltan 300
QY 301 QHSDVICQSEPDSPSSGVSLSYEVERCOOLSATILTDHOYLERTPLCAILKQKAPQY 360
DB 301 qhsdvlcseqpddsfpsgsvslyevercqolsatiltldhoylerltpcailkqkappqy 360
```

RT orf1, orf2, orf3, orf4 (DNA segment VD1) [parvo-like virus, Yamashita
RT isolate, host-silkworm, Genomic, 6542 nt].
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB033596; BA85362.1; -
KM Nonstructural protein.
SO SEQUENCE 1115 AA; 128304 MW; 7C0DA2DCFA746428 CRC64;

Query Match 4.2%; Score 115; DB 12; Length 1115;
Best Local Similarity 19.8%; Pred. No. 0.79;
Matches 122; Conservative 93; Mismatches 186; Indels 216; Gaps 35;

OY 5 PATNYITP-----LNO-----LKGITV-----NYGVVFKFPYLSKGTDCSV 46
DB 363 PAENFIFTKSVLNLDEVEFNDEGEIIFDNIEMLTDLYSIIKKFOI-----NDYGKE 416
OY 47 VTIVDQTNVNLTC-----LFSGVYELPIIYKNGDIVRFRHLK- 85
DB 417 ANLIDGGERKCLACKNNKKGYVARHGLGIFGYCRSCFLAHNNTPIVYFNFKGY 476
OY 86 -----IOVYKRETOGITSSG-----FASLTREGTLGAPIIPRTSSKY 122
DB 477 DHIHLDELKLNKSKHNCRGKSIKMDVITHKDLSDFIITFKDI----- 523
OY 123 FNFTEDHKVYALRWASTHMSPTLLKLCV-----QPMOYFDLTCOLLGKAEV 174
DB 524 FNF-----LPESLASLANKLTLTKYTPDKFKDAFNSGKEFPYEMED-DFNKLEIEVP 576
OY 175 -----DGASFLK-----VW-DGTRTFPSMRVLIOD-----LVLE----- 204
DB 577 ODPADMSRLTNKKGTETIIKKANOINIDNNMOIFHEVLYLNLDELWMLLEVEAFRTD 636
OY 205 -----GDLSHIRLQNLTDI-LVYDNVH--VARSLSKGSFLRIYSLH 245
DB 637 TVNEKDIDPVYFDGAPGLTFYLARMYENSLDMHYIPDKNYLDVSRIRVG-----VTQVV 692
OY 246 TKLOSMSSENOTMLSEFLHSGTSGRGIRVLPESNSDVQLKKDESAULTANQMSDV 305
DB 693 TKYANIEDVDETIYLLD--VNTMYSTCKOKLANKYLGTDITLTDNDVS-----DDN 742
OY 306 ICO-SEPDPSFGSSGVSYLEVERCOOLSATI--LTDHOYLERTPLCALIKOKAPOYR 361
DB 743 FCYIYKGDFTSP-----EYLHDLPAHLSMPLMHQV--NNKLCITFLDK--KDL 788
OY 362 IRAWLSRKYPRRLFOYKYLHCPKCHLQOEVPH-----GDLDIIFODGATK--TPD-VK 412
DB 789 IHSKVFYV-----YLSKGLVCDKTHVYKFKQEVYIKDYVEINIOKRNSSIDPGIKDYK 843
OY 413 LQNTSLVDS-----KIMTKNOKG-----RKVAVHFVNNGLPLSNCL--LL 454
DB 844 LKNNALGCKTCENFKYKIFSVTVNNSGDRENKCMKSAKSHITLGNCL--LYECVIRYL 901
OY 455 IE-----GGTLEITCKL 466
DB 902 LDKPIQIGFTLLELAKL 918

RESULT 12
OY9C2K8 PRELIMINARY; PRT; 759 AA.
AC OY9C2K8;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CONSERVED HYPOTHEITICAL PROTEIN.
GN 3410.40.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OX Sordariales; Sordariaceae; Neurospora.
OX NCBI_Taxid-5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Algen V., Hoheisel J., Brandt P., Fartmann B., Holland R.,

RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL513442; CAC28643.1; -
SO SEQUENCE 759 AA; 85218 MW; DC737A5574FF4F73 CRC64;

Query Match 4.1%; Score 111; DB 3; Length 759;
Best Local Similarity 20.7%; Pred. No. 0.99;
Matches 63; Conservative 50; Mismatches 125; Indels 66; Gaps 12;

OY 3 LVPTNIVYIRPLNQL-----KGGTIVNYGVYFKFPYLSKGTDCSVYTI-----VD 51
DB 18 LFEST---LTPRLALIDDGDAQPGSNVNYIGLDCRAVAVHGSQKCTLTISLSTIE 74
OY 52 QTNVNLTCFLSGVYELPIIYKNGDIVRFRHLKIQYKRETOGITSSGFASLTREGTLG 111
DB 75 DESAGVELVIF--RPEARMPVEGAGDVLVLSAKVQRFKSNPKSLITSKITTCVYKKAFT 132
OY 112 APII-----PRTSKYFNFTEDHKVYAL-----RYMASTHM 144
DB 133 IPVYPASAOYALLPPKQGESHKLLKEEHQYVSYLNVYIDKYDVPDEAYQQRVYKSLNV 192
OY 145 SPSTVTLKLDVOPMOYFDLTCOL-----LQKAEV-----DGASFLKWDGTRTP 190
DB 193 KDRFSLLK--DIVDGNRYDLIGQVAKPDYDEMKRILYLSQDTENDLFHHYMEGRDL 250
OY 191 FPSNR---VLIODLVLEGLSH--IHRLONLTDILVYDNVHVAR--SLKVSFLRIYS 243
DB 251 ASAARPADAYVEDNNPMAOOHPVGVYGRRTIOISCYDAHADFIIRAGVSAGMMLSLRN 310
OY 244 LHTK 247
DB 311 VOYK 314

RESULT 13
OY9XE80 PRELIMINARY; PRT; 483 AA.
AC OY9XE80;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE SERINE-TYPE CARBOXYPEPTIDASE.
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Sorghum.
OX NCBI_Taxid-4558;
RN [1]
RP SEQUENCE FROM N.A.
RA Liaca V., Lou A., Messing J.W.;
RT "Microsytieny analysis of 22-kda zein cluster in maize and sorghum."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF061282; AAD22150.1; -
DR HSSP: P08819; 1WHT.
DR InterPro: IPR000379; Est_1ip_thioest_actsite.
DR InterPro: IPR001563; Serine_carptsept.
DR Pfam: PF00450; serine_carptsept. 1.
DR PRINTS: PR00724; CRBOXYPTASEC.
DR PROSITE: PS00131; CARBOXYPEPT_SER; UNKNOWN_1.
KM Carboxypeptidase.
SO SEQUENCE 483 AA; 53413 MW; DB91DCGEF772A0D9 CRC64;

Query Match 4.0%; Score 110.5; DB 10; Length 483;
Best Local Similarity 20.6%; Pred. No. 0.57;
Matches 91; Conservative 69; Mismatches 134; Indels 147; Gaps 28;
OY 95 GITSSGFASLTREGTLGAPIIPRTSSKYFNFTEDHKVYALRWASTHMSPTLLKLC 154

Matches 109; Conservative 83; Mismatches 177; Indels 103; Gaps 23;

```

0Y 18 KGGTIVNYYGVVKKFKRPPYLS-KGT-DYCSVYTVDO-----ANKLTCLLFSGNYELP1 71
Db 38 KKNITVLEFVIGVIOFTBPSROSLHGTIKOMVITVYVLMPTCDTSSIGLOJHLFSCQNDLPV 97
0Y 72 IYKMGDIYFRHRLKIOVYKKTGOTITSSGFA-----SLTFEGTCLAP1IPRTS----- 119
Db 98 IKOVGPPLLHQITLRSYRDRGTQGLSKDQFRYALMPPFSSMSKOTLCRQPMPLMTGDK 157
0Y 120 -----SKFNTTTEHKKVLELRYWAS-----THMS-PSWILLKACDVOPMC-----FDLTC 166
Db 158 EEOFALLINKIMDEOTNKKHKGELLSTSSAQNOTGLSTYSPVSSLQITPHQSCFYA 217
0Y 167 QLLGKAEVYDGAFLKLKVDGTR-----TPFPSS-WRVYLIOULVEGSLSHIRLQNL 216
Db 218 QVI-KTWYSOKNTFLVYDIYENLEFPMSYTSSSMR-----GFFG-----RF 261
0Y 217 TIDILYVDNHHVAVAS-LKVGSEFLRIYSLHTKLOSMSNENOTMLSEFHLHGTS-YGR 273
Db 262 SIRCLMEHDFYCSNVIKEGDYVVMNVKRIKIDHLG-----YLECIIJHGSKARYNM 314
0Y 274 GIRVLPESNSVDOLK-KDLESANLTRANQSDVIGS-----EPDDEPSSGSUSLY 324
Db 315 SIEKVDSEEPFLNETKSKRRLYON-----CONGLEAVIEKUSQOSQSENPIAH 364
0Y 335 EVER-COOLSAITLTDHQLERTPLCALIKOR-----AFQOYRIAKLUSYKFRRLFGS 377
Db 365 ELKQTSVEIITAHVINEPASLKLITISILHAPLONLLKPRKRLRYQVVDWFMKSLTQF 424
0Y 378 VKHGRPKH-----LLOEYPRHGGDIIF-QDGAKTRPVKLOMTSLYD 420
Db 425 AVLSOPBSSVYMFALLVYRDSVN-TLPVJFEPDSDAEELINSSKIQCNLAD 475

```

RESULT 7		PRELIMINARY:		PRT:	463 AA.
09FNH7					
AC	09FNH7;				
DT	01-MAR-2001 (TREMBLrel. 16, Created)				
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)				
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)				
DE	GB A029059.1.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Euarabota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:				
OC	Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:				
OC	eucosids. II: Brassicales: Brassicaceae; Arabidopsis.				
OX	NCBI_Taxid-3702:				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-COLUMBIA:				
RX	MEDLINE-98069011; PubMed-9405937;				
RA	Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,				
RA	Tabata S.;				
RT	Structural analysis of Arabidopsis thaliana chromosome 5. II.				
RT	Sequence features of the regions of 1,044,062 bp covered by thirteen				
RT	physically assigned pl clones."				
RL	DNA Res. 4:291-300(1997).				
DR	EMBL, AB006700; BAB08953.1.				
Q0	SEQUENCE 463 AA; 53630 MW; A735DBF109155D7D CRC64;				

Query Match	5.68;	Score 153.5;	DB 10;	Length 443;
Best Local Similarity	22.88;	Pred. No. 0.00012;		
Matches 94;	Conservative 59;	Mismatches 164;	Indels 55;	Gaps 17;

Oy 23 VNYGVYKFRPPLTSGIDCVSVTVD--QTWVKTLCLFSCNTEALPIIKKNCIVR 80
|| : : || :
Db 25 VNLIQVIELG---FSNGSDCSCTLKIDVPYSSGLPKVFATINDLPVES:GDIL 81

Oy 81 FHRKLIOYKKETOGI-----TSGSASLTFTGICADIPRTSKSYNTFTDHKNVEAL 136
|| : : || :
Db 82 LSRPAKIYLINRKITALCNETTSSFA--LNGKHVSVDIPVSSPKFLMEQRNF:SNL 139

```

0Y 137 RWMASTHM-----SPSMITLKLCDVOPMOPYFPLDTOLGKAAVDASLTKVMOSTRP-- 190
Db 140 REMATTYAFEDGSCCFTSLK -DIKEGECSNLSCDIYHISVYDRNMTLPMOSTEMPPC 197
0Y 191 -----FPMW----PVLIDVLEGGDLSHRLNLTIDIL 223
Db 198 NILVSERLPLCEPEEMLTMYLAKFPFGSYLRIIT-DVSEKOA--IHOLG----- 248
0Y 222 YVDNKHVAR--SLKVGSLRLIYSLHKILO-SNMSENOTMLSLIEPHLHGSTGRCIRY 277
Db 249 ---GQHVKLNLNFYOYVNMGLMNNATFPSTKMQOYTSRENEAFSPORFLTPPLIMCEKF 306
0Y 278 LPESNSDVOLKRLDESANLTANOHSDVICOSEPDPSEPPSSGSVSLYEVCQOLSATIL 337
Db 307 SFRMN-----PIARCSIRSHSEITGVAHDAPE-----VSLMDLITLYHNVTAKFR 351
0Y 338 TDHXYLETRP-----ICALLKQAKAOQYIRAKLASYPKRLPF 375
Db 352 CVYFRIOYPRDYKRLDINGNIKIVALEDATA--RIHSLVADGECKF 400

```

```

RESULT      8
076380
ID      076380      PRELIMINARY:      PRT:      497 AA.
AC      076380.
DT      01-NOV-1998 (TREMBLrel. 08, Created)
DT      01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE      ALPHA TELOMERE BINDING PROTEIN.
OS      Oxytricha trifallax.
OC      Eukaryota: Alveolata: Ciliophora: hypotrichs: Slicotrichida:
OC      Oxytrichidae: Oxytricha.
OX      NCBI_TaxID=5946;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Prescott J.D., Dubois M.L., Prescott D.M.:
RT      "Oxytricha trifallax macronuclear alpha telomere binding protein
RT      gene."
RL      Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
DR      HMLL: AF067831; AAC27615.1; -.
DR      HSSP: P29549; IOTC.
DR      InterPro: IPR003415; Telo_bind_alpha.
DR      Pfam: PF02307; Telo_bind_alpha; 1.
SQ      SEQUENCE 497 AA: 56361 MW: 132C768DC3AB686 CRC64:

```

Query Match	5.3%	Score 144.5;	DB 5;	Length 497;
Best Local Similarity	20.6%	Pred. No. 0.00078;		
Matches	87;	Conservative	67;	Mismatches 156;
			Indels	113;
			Gaps	18;

```

0Y 7 TNYITPLN--OLKOGSTLVNYGVYKFFKPPYLSKGDYGVNIYD-----Q 52
Db 34 SKYEVELTKAQNOLSVTAQNHFRAYVIAIDATFFYKTIQNOEXICSLKIVDPSLTKREKGTG 93
0Y 53 TNVKLTCLLFGSGNTEALPIIYKNDIYVFKHLKIQYKKEIQG-----ITSGRPSLTF-- 106
Db 94 NSDIATVLVYKREDELPIIHRGLDIIHIAHATIRLINOGEQFNANIFYSSSMALFEJTDK 153
0Y 107 -----EGTLCAPLIIPRT--SSKYFNFTTEDHKWEALRVNASTHM-----SPSW 148
Db 154 KSALOELIGOGAET--SDLYVFPESHSGKNFTFOKNENAGIYONIRKMAAOOYFTOYAVISNDMF 211
0Y 149 TLLKLCIDVQPMQYFDLTCLLQKGAKEVDASFLKLVWMDOTRPPFSWBYLJODIYLEDLS 208
Db 212 TALNKAQOKKD--FDYVAKIIQIHEDDEYITELKAKDASGVF-----YTLLAKIKFP 263
0Y 209 HIH-----RLONLITD-----ILYDNNHVVARSLSKVSFLAIYLSLHKIKLSMSENQ 256
Db 264 HLRTGEVVRIRSAIYDETSTOKWALLLSHYSNIYVTFISTKLAA--ELRGHITDDRAVEK 321
0Y 257 TMLSLFHL-----HGGTSYGRGIVALE-----SNSVDOLKIKDLESANLT 298

```

```
Db 655 LEEVHHSLGCCSKYIGLYLNDOCKSLKRP-CVNWSELEYLGLRSCDSLEKLPETIYGRMKP 713
OY 365 KLRSTK-----PRRLFOSVKLHCPK-----CHL 387
Db 714 EYIHMGGSGIRELPSSIFQ-YKTHVTKLLMNMKNLVALPSSICRLKSLVSLVSGCSK 772
OY 388 LQEVPRH-GDLDIIFQDGATKTPDVKIQNTSLYDSKIWTIKNKGKRVAVHF-----438
Db 773 LESLPEEIGDLDNLRVFDASDTLLLRPPSSIRLNKL-IIMFRGFKDGVHFEFPPVAEG 831
OY 439 YKNNGILPLSNCELLIEGTLSEICLSNKFNSYIPVRSCHDELIDLS 489
Db 832 LHSLEYLNL--YCNLIDGGLPEEIGSLSS-----LKKLDES 865
```

Search completed: April 9, 2002, 17:08:56
Job time: 275 sec

```

STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LVZNMOT01
CLONE: 348429
US-09-274-570-3

```

```

Query Match      3.3%; Score 91; DB 3; Length 425;
Best Local Similarity 19.5%; Pred. No. 0.96;
Matches 92; Conservative 81; Mismatches 175; Indels 124; Gaps 22;

```

```

OY 23 VAVYGVAFKFP--YLSKGD--YCSVYIVDOINVKLTCL---LFGSGNEALPIIKN 75
DB 16 LSLSGHVGFDLSLPDOLVNRKSTSGCFNLGCGETGICKSTLMDLTFNKFESDEPATINE 75
OY 76 GDVIFHRLKIOVKRKEGTGISTSGEASLTFEGTIG-----APIIPRTSKYFN 124
DB 76 PGV-----RLKARSTYLOESNR-----LKLTIYIVGFGDQINKDSYKFIYEVIOAOFEA 127
OY 125 FTEDHKRVEAL-----RYMASTH--MSPSWTLKRLCDVOPQYFDLTCOLLGKAEVDGA 177
DB 128 YLQGEELKIKRSLFNHYDRIHACLYFIAPTGHSLKSLDLVTKKLDG----- 174
OY 178 SFLKAVMGRTPRPSWVLIODLVLEGD---SHIHLQNLITDILYDNHVAHARSLK 234
DB 175 -----RVNIIPIIAKADTIAKNELHFKSKIMSELSVNG----- 208
OY 235 VGSFLRIYSLTKLOSMNSENOTM--LSLEFHLGCT-----SYGRG--IRVLP 279
DB 209 ---VQIOTFIDEETVAENATMSVHLPLFAVVGSTEEKIKGNKAKARQYRWGVQVEN 264
OY 280 ESNVDOLKDLKDESANL-----TANQSDVI--COSEP---DDSPSSGVSCL---YE 325
DB 265 ENHCDFVLRRLIYVNNEDLREQTHRTHYELRYCKLEEMGFKDTDDSKFFSLQETYE 324
OY 326 VERCOOLSATITDHOYLERTPCALILKOKAPO---OYRIKARLSYKPRRLFQSVK 381
DB 325 AKRNFLE--LQKKEEHROMFWRKVEKEALKEAELEHKEKDLKRTHOEKKRVE 382
OY 382 CPKCHLOEVPHEGDLIIFODGATKTPDVKLONTSLYDSKIWTITKNOKGR 433
DB 383 DKRKELEEVNN-----FQ---KKRAAOQLQSOAQOQKQKDKK 424

```

```

RESULT 13
US-08-323-1708-2
Sequence 2, Application US/083231708
Patent No. 5733772
GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
APPLICANT: Kaslow, David C.
TITLE OF INVENTION: Cloning and Expression of Plasmodium
TITLE OF INVENTION: Falicparum Transmision-Blocking Target Antigen, Pfs230
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,1708
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409

```

```

FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-113100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3135 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-323-1708-2

```

```

Query Match      3.3%; Score 91; DB 1; Length 3135;
Best Local Similarity 18.3%; Pred. No. 29;
Matches 105; Conservative 89; Mismatches 185; Indels 194; Gaps 28;

```

```

OY 20 GIIVN---VGV---VREFFKPPYLSKGTDY-----CSVYTIYDQTN-VKLTCLFSGN 65
DB 825 GNLVNSVYVYKEMNAKFNQYVHIPISTYDITLNLFCIIKKEESNLSTLYVYSI 884
OY 66 YEALPIIKNGDIYRFH-----RLKIOYRK-----ETOGITSSG- 100
DB 885 NEEL-----NFSLFDFESFVPIKKTIOVAQRNVNKEHDYTCDFDKDKTVPSTANGK 939
OY 101 -----FASLTFEGTL-----GAPLIIPRT-----SSKYFN 125
DB 940 KLFICRKLKEEDFTTLCKVANKTOYPIEIPKTLKDKKVEVLKLDIDIOYOMFSKFFK 999
OY 126 TTEDHKRVEA---LRMASTH-----SPSWTLKRLCDVOPQYFDLTCOLLGKAEVD 175
DB 1000 NTQNAKYLVLPYLIIFPNHIGKELKNPTK-----NKKDVKYEQSSVLSPLSSAD 1054
OY 176 GASFLKAWDGRTPFPSPRWLIODLVLEGDLSHIHLQNLITDILYDNHVAHARSLK 235
DB 1055 SLGKLNLFDLOET---VCLTEKI-----RYLNLSINELGSDNNT--FSVYFV 1098
OY 236 GSFLRIYSLTKLOSMNSENOTMLSLEFHLGCTSYGRGIRVPESSDDVQD----- 288
DB 1099 PPIYDIKE-----PFYFMFGCANNKG-----EGNIGIVELLISKOE 1135
OY 289 -----KKDESANLITANQH-----SDVICSEBDDSPSSGSVSYLEVERCO 330
DB 1136 KIKGCFHESKIDYFNENISDTECHLHAYENDIIGFNCLETHHPNEVEVEYEDAEIYL 1195
OY 331 OLSATITDHOYLERTPCALILKOKAPOQYRIKAK-----LSYKPRRLFQSVKIKHCKCH 386
DB 1196 QPENCFNNAVYGLNSVDITTLKN--AOTYININKKTPFLKIPPNLLEDEVEISC-OCT 1252
OY 387 LLOEVPHEGDLIIFODGATKTPDVKLONTSLYDSKIWTITKNOKGRVAVAFVANNILP 446
DB 1253 IKQYVK---KIKVITKNDYVLKREVQSESTLDDKIKKCHE-----NINP 1257
OY 447 LSNECLLIEGTLSE---IC--KLSKRENSV 473
DB 1298 RVNK-----TFDENVEYTCNIKIEFENFYI 1322

```

```

RESULT 14
US-08-310-912A-108
Sequence 108, Application US/08310912A
Patent No. 5981730
GENERAL INFORMATION:
APPLICANT: Ausubel, Frederick M.
APPLICANT: Staskavicz, Brian J.
APPLICANT: Brent, Andrew F.
APPLICANT: Dahlbeck, Douglas
APPLICANT: Katagiri, Fumaki
APPLICANT: Kunzel, Barbara N.

```


Query Match	3.4%:	Score 93.5:	DB 2:	Length 1451:
Best Local Similarity	21.3%:	Pred. No. 4.4:		
Matches	76:	Conservative	51:	Mismatches 118:
			Indels	111:
			Gaps	19:

Query Match	3.3%	Score 91	DB 2	Length 425
Best Local Similarity	19.5%	Pred. No.	0.96	
Matches 92	Conservative 81	Mismatches 175	Indels 124	Gaps 22

175 -----KVNIPIAKADTIAKNEIHKFKSKIMSELSVNG----- 208

ATTORNEY/AGENT INFORMATION:
 NAME: Henry, Janis C
 REGISTRATION NUMBER: 34,347
 REFERENCE/DOCKET NUMBER: 2631
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206)470-4189
 TELEFAX: (206)233-0644
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1569 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-09-458-791-2

Query Match 3.5% Score 94.5; DB 4; Length 1568;

Best Local Similarity 20.0%; Pred. No. 3.9; Mismatches 164; Indels 171; Gaps 26;

Matches 99; Conservative 60; Mismatches 164; Indels 171; Gaps 26;

QY 91 KETGGIT-----SSGFSALTF-----EGTLCAPLIPTSSKYENFTEDHKMV 133
 DB 208 KDTGRLATQELGRKLCBAGSLHFVDAFLMNGSIYFPY-----YNTS----- 255
 QY 134 EALVMASTHMSPSWTLKLCDDVPMQYFDLTCLLGAEDVGSFLK-----VMD 185
 DB 256 -----GAATGMPMARIAOSTEVLFGQASLDC---GHGHPDGRRLSSSLVEALDYMA 307
 QY 186 GT-----RTPEPSMVLIDLVLEGLSHIH-RLQNLITIDILVYDNHV-----HV 229
 DB 308 GVFSMAAGEGERSPTTALCLFR-----MSEIQARAKRVSMDFTAESHCKEGDOP 360
 QY 230 ARSLKVSFLRYSLHTRKLOSMNSENOTMLSLFEHLHGTSYGCGRIVLPESNEDVDOLK 289
 DB 361 ERVQPIASSTLIHSDLSVYGVVNMRTVFL-----GTGDG-----QLL 400
 QY 290 KDESANLTANQHSVDVLCQSEPDSPSSGSVSLYEVEKCOQLSATILTDHQLERTPL- 348
 DB 401 KVIIGENLTSN-----C---PE-----VITEIK-----EETPVF 426
 QY 349 -----CALIKOKAPOQYRIR-AKLRSYKP-RLIFOSVKLHCPKCHLLOEVPHEGD 396
 DB 427 YKLVPDPVKNIYIYLLAGKEVRIRIVANCNKHKSCSECLTATDPHCGCHSLQCTFOGD 486
 QY 397 -----LDITFQCATKTPPVKCLONTSLYDSKITTTNOKGRKVAHVFAKNGILP 446
 DB 487 CVHSENLLENWLDI--SSGAKKCPKIQI-----IRSKKERTVTYMG-----SFS 529
 QY 447 LSNCELLIEGTLSEICKLSNKFNSV-----IPVRSCHEDLELDLAPFLIOGTIHHYG 502
 DB 530 RHSCGMKRV-NVDSRELCONKSOPTNRCTCSIPTRATIKDVSVMVWFSGSNLSDRFN 588
 QY 503 TGYCTP-----PICV 512
 DB 589 FTNCSLKECPACV 602

RESULT 6
 US-09-459-066-2
 : Sequence 2, Application US/09459066
 : Patent No. 6187909
 : GENERAL INFORMATION:
 : APPLICANT: Spriggs, Melanie
 : TITLE OF INVENTION: VIRAL ENCODED SEMAPHORIN PROTEIN
 : TITLE OF INVENTION: RECEPTOR DNA AND POLYPEPTIDES
 : NUMBER OF SEQUENCES: 10
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSEE: Janis C. Henry
 : STREET: 51 University St.
 : CITY: Seattle
 : STATE: WA
 : COUNTRY: US

ZIP: 98101
 COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: MS-DOS/Windows 95

SOFTWARE: Word for Windows 95, 7.0a

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/459,066

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/958,598

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Henry, Janis C

REGISTRATION NUMBER: 34,347

REFERENCE/DOCKET NUMBER: 2631

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206)470-4189

TELEFAX: (206)233-0644

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1568 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-459-066-2

Query Match 3.5% Score 94.5; DB 4; Length 1568;

Best Local Similarity 20.0%; Pred. No. 3.9; Mismatches 164; Indels 171; Gaps 26;

Matches 99; Conservative 60; Mismatches 164; Indels 171; Gaps 26;

QY 91 KETGGIT-----SSGFSALTF-----EGTLCAPLIPTSSKYENFTEDHKMV 133
 DB 208 KDTGRLATQELGRKLCBAGSLHFVDAFLMNGSIYFPY-----YNTS----- 255
 QY 134 EALVMASTHMSPSWTLKLCDDVPMQYFDLTCLLGAEDVGSFLK-----VMD 185
 DB 256 -----GAATGMPMARIAOSTEVLFGQASLDC---GHGHPDGRRLSSSLVEALDYMA 307
 QY 186 GT-----RTPEPSMVLIDLVLEGLSHIH-RLQNLITIDILVYDNHV-----HV 229
 DB 308 GVFSMAAGEGERSPTTALCLFR-----MSEIQARAKRVSMDFTAESHCKEGDOP 360
 QY 230 ARSLKVSFLRYSLHTRKLOSMNSENOTMLSLFEHLHGTSYGCGRIVLPESNEDVDOLK 289
 DB 361 ERVQPIASSTLIHSDLSVYGVVNMRTVFL-----GTGDG-----QLL 400
 QY 290 KDESANLTANQHSVDVLCQSEPDSPSSGSVSLYEVEKCOQLSATILTDHQLERTPL- 348
 DB 401 KVIIGENLTSN-----C---PE-----VITEIK-----EETPVF 426
 QY 349 -----CALIKOKAPOQYRIR-AKLRSYKP-RLIFOSVKLHCPKCHLLOEVPHEGD 396
 DB 427 YKLVPDPVKNIYIYLLAGKEVRIRIVANCNKHKSCSECLTATDPHCGCHSLQCTFOGD 486
 QY 397 -----LDITFQCATKTPPVKCLONTSLYDSKITTTNOKGRKVAHVFAKNGILP 446
 DB 487 CVHSENLLENWLDI--SSGAKKCPKIQI-----IRSKKERTVTYMG-----SFS 529
 QY 447 LSNCELLIEGTLSEICKLSNKFNSV-----IPVRSCHEDLELDLAPFLIOGTIHHYG 502
 DB 530 RHSCGMKRV-NVDSRELCONKSOPTNRCTCSIPTRATIKDVSVMVWFSGSNLSDRFN 588
 QY 503 TGYCTP-----PICV 512
 DB 589 FTNCSLKECPACV 602

RESULT 7
 US-09-308-375-2
 : Sequence 2, Application US/09308375

```

Db 939 SNEYEKREALKVTHLNIIGECNIQFGINPOTG---EXCII-----EVNARLSRSA 990
OY 62 FSGNTEALPIIYKNGDIVRFH---RLKIQVYKKEGOTITSSGFASLTFEGILG--APIIP 116
Db 991 LASRATGYPYIAISAKIALGDLISLKNSTTKTT-----ACPEFSLDYITTKIP 1040
OY 117 RTSSKYENFTT---EDHKWVEALRWMASTHMSPSWTLKLCVDQVPOQYFDLTCLLGKA 172
Db 1041 RMDLKNKFEFASNTMNSKSKSGEVMISIGRTFEESIQRCDIDNNYLGFSNTYC----- 1094
OY 173 EVDGASFLKAWDGTTRTFPSMRV--LIQDLVLEGDSLHRLQNLITDILVYDNHVA 230
Db 1095 -IDWDE--KRIIEELKNPSK-RIDAIHQAFHLMNPMKIHLELHIDWFL----- 1141
OY 231 RSLKGSFLRIYSLHTRKLSNSENQOTMLSEFHLHGTSYGRGIRVLPESNSVDOLK 290
Db 1142 -----HKFYNIYNLQNKLTLEQLSFNDLKYFKKHGFSQKQIAHYLSFTSDNNNNN 1196
OY 291 DLESANLTANOHSVDYICQSPDPSFGSSGVSLEYVERCOOLSATI--LIDHOYL----- 343
Db 1197 NISSCRVTEN---DVMKYREKLGLEP-----HIKVIDLSAEFPALITNYLYLYOQO 1245
OY 344 ERTPLCALIKOKAPOOYIRAKLRBSYKPRRLFQSVKLCPCCHLQEVPHGEDLDIIFOD 403
Db 1246 EHDVLPYLNKKR-----KICTLNKRNANKKKVHV-KNHLVNEVVDKQDOLHKN 1295
OY 404 GATKTPDKLQNTSLYDSKIWTTRKNQGRKVAHVFKNNGILPLSNECLLIEGGTISEI 463
Db 1296 NNNNN-----MNSGVNEKCKLNKESYGYNNSSNCINTNINNIENNIC-----HDISINKN 1346
OY 464 CK--LSNKFNSYIPVRSGHEDLELDLSAPFLIOGTIIHYG 502
Db 1347 IKVTINNSNISI-----SNNEVE-TNLNCVSEBAGSHIYG 1382

```

RESULT 2

```

US-09-150-741-2
; Sequence 2, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2391
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2

```

Query Match 3.98; Score 106; DB 4; Length 2351;
 Best Local Similarity 20.98; Pred. No. 0.52;
 Matches 109; Conservative 80; Mismatches 230; Indels 102; Gaps 25;

```

OY 7 TINTIYPLNG--LKGGTIVWYGV--VKFKPPYLSKGTIDYCSVITVDOINXLI--CLL 51
Db 939 SNEYEKREALKVTHLNIIGECNIQFGINPOTG---EXCII-----EVNARLSRSA 990
OY 62 FSGNTEALPIIYKNGDIVRFH---RLKIQVYKKEGOTITSSGFASLTFEGILG--APIIP 116

```

```

Db 991 LASRATGYPYIAISAKIALGDLISLKNSTTKTT-----ACPEFSLDYITTKIP 1040
OY 117 RTSSKYENFTT---EDHKWVEALRWMASTHMSPSWTLKLCVDQVPOQYFDLTCLLGKA 172
Db 1041 RMDLKNKFEFASNTMNSKSKSGEVMISIGRTFEESIQRCDIDNNYLGFSNTYC----- 1094
OY 173 EVDGASFLKAWDGTTRTFPSMRV--LIQDLVLEGDSLHRLQNLITDILVYDNHVA 230
Db 1095 -IDWDE--KRIIEELKNPSK-RIDAIHQAFHLMNPMKIHLELHIDWFL----- 1141
OY 231 RSLKGSFLRIYSLHTRKLSNSENQOTMLSEFHLHGTSYGRGIRVLPESNSVDOLK 290
Db 1142 -----HKFYNIYNLQNKLTLEQLSFNDLKYFKKHGFSQKQIAHYLSFTSDNNNNN 1196
OY 291 DLESANLTANOHSVDYICQSPDPSFGSSGVSLEYVERCOOLSATI--LIDHOYL----- 343
Db 1197 NISSCRVTEN---DVMKYREKLGLEP-----HIKVIDLSAEFPALITNYLYLYOQO 1245
OY 344 ERTPLCALIKOKAPOOYIRAKLRBSYKPRRLFQSVKLCPCCHLQEVPHGEDLDIIFOD 403
Db 1246 EHDVLPYLNKKR-----KICTLNKRNANKKKVHV-KNHLVNEVVDKQDOLHKN 1295
OY 404 GATKTPDKLQNTSLYDSKIWTTRKNQGRKVAHVFKNNGILPLSNECLLIEGGTISEI 463
Db 1296 NNNNN-----MNSGVNEKCKLNKESYGYNNSSNCINTNINNIENNIC-----HDISINKN 1346
OY 464 CK--LSNKFNSYIPVRSGHEDLELDLSAPFLIOGTIIHYG 502
Db 1347 IKVTINNSNISI-----SNNEVE-TNLNCVSEBAGSHIYG 1382

```

RESULT 3

```

US-08-264-002-2
; Sequence 2, Application US/08264002
; Patent No. 5559019
; GENERAL INFORMATION:
; APPLICANT: GUI, XIANG-FANG
; APPLICANT: FU, XIANG-DONG
; TITLE OF INVENTION: NOVEL PROTEIN SERINE KINASE, SRPK1
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
; STREET: 1880 Century Park East, Fifth Floor
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/264,002
; FILING DATE: 22-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: TUMARKIN PH.D., LISA A.
; REGISTRATION NUMBER: P-38,347
; REFERENCE/DOCKET NUMBER: PD5590
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/455-5100
; TELEFAX: 619/455-5110
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 655 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-264-002-2

```

ID	NRPI_XENLA	STANDARD:	PRT:	928 AA.
AC	P28824:			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	NEUROFILIN-1 PRECURSOR (A5 PROTEIN) (A5 ANTIGEN).			
OS	Xenopus laevis (African clawed frog).			
OC	Eumetazoa: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:			
OC	Amphibia: Batrachia: Anura: Mesobatrachia: Pipidoidea: Pipidae:			
OC	Xenopodinae: Xenopus.			
OX	NCBI_Taxid=8355;			
RP	SEQUENCE FROM N.A.			
RP	TISSUE=Brain;			
RC	MEDLINE=9137458; PubMed=1908252;			
RA	Tagaki S., Hirata T., Agata K., Mochii M., Eguchi G., Fujisawa H.;			
R1	The A5 antigen, a candidate for the neuronal recognition molecule,			
R1	has homologies to complement components and coagulation factors.*;			
RL	Neuron 7:295-307(1991).			
CC	-I- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE			
CC	CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF			
CC	CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS			
CC	SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY			
CC	SIMILARITY). PRESUMED TO BE INVOLVED IN THE NEURONAL RECOGNITION			
CC	BETWEEN THE OPTIC NERVE FIBERS AND THE VISUAL CENTERS.			
CC	-I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			
CC	-I- TISSUE SPECIFICITY: RETINAL GANGLION CELLS AND VISUAL CENTER			
CC	NEURONS.			
CC	-I- SIMILARITY: BELONGS TO THE NEUROFILIN FAMILY.			
CC	-I- SIMILARITY: CONTAINS 2 CUB DOMAINS.			
CC	-I- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.			
CC	-I- SIMILARITY: CONTAINS 1 MAM DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: D10467; BAA01260.1; -			
DR	InterPro: IPR000859; CUB.			
DR	InterPro: IPR000421; FA58_C.			
DR	InterPro: IPR000998; MAM.			
DR	pfam: PF00431; CUB: 2.			
DR	pfam: PF00754; F5_FB_Type_C: 2.			
DR	pfam: PF00629; MAM: 1.			
DR	PRINTS: PR00020; MAMDOMAIN.			
DR	SMART: SM00042; CUB: 2.			
DR	SMART: SM00231; FA58C: 2.			
DR	SMART: SM00137; MAM: 1.			
DR	PROSITE: PS00740; MAM_1: 1.			
DR	PROSITE: PS01180; CUB: 2.			
DR	PROSITE: PS01285; FA58C_1: 2.			
DR	PROSITE: PS01286; FA58C_2: 2.			
DR	PROSITE: PS50060; MAM_2: 1.			
KW	Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;			
KW	Antigen.			
FT	SIGNAL	1	21	POTENTIAL.
FT	CHAIN	22	928	NEUROFILIN-1.
FT	DOMAIN	22	860	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	861	883	POTENTIAL.
FT	DOMAIN	884	928	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	27	141	CUB 1.
FT	DOMAIN	147	265	CUB 2.
FT	DOMAIN	275	424	F5/8 TYPE C 1.
FT	DOMAIN	431	584	F5/8 TYPE C 2.
FT	DOMAIN	646	812	MAM.
FT	DISULFID	27	54	PROBABLE.
FT	DISULFID	82	104	PROBABLE.
FT	DISULFID	147	173	PROBABLE.
FT	DISULFID	206	228	PROBABLE.

FT	DISULFID	275	424	BY SIMILARITY.
FT	DISULFID	431	584	BY SIMILARITY.
FT	CARBOHYD	150	150	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	261	261	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	300	300	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	523	523	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	844	844	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	SEQUENCE	928 AA:	104416 MW:	AF6B323B0A4C789D CRC64:

Query Match	3.6%	Score 97.5;	DB 1;	Length 928;
Best Local Similarity	18.9%	Pred. No. 12;		
Matches	91;	Conservative	69;	Mismatches 186;
				Indels 135;
				Gaps 22.;

[illegible]

Search completed: April 9, 2002, 17:11:19
Job time: 248 sec

CC CHANGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
 CC PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
 CC SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
 CC SUBSTRATE FOR CELLS.
 CC -1 SUBUNIT: PURIFIED PERFECTAN HAS A STRONG TENDENCY TO AGGREGATE IN
 CC DIMERS OR STELLATE STRUCTURES.
 CC -1 SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1 TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
 CC -1 PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
 CC AND O-LINKED OLIGOSACCHARIDES.
 CC -1 SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1 SIMILARITY: CONTAINS 10-5 LAMININ EGF-LIKE DOMAINS.
 CC -1 SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
 CC -1 SIMILARITY: CONTAINS 22 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1 SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
 CC -1 SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
 CC -1 SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M62515; CAA4373.1; -
 CC EMBL: M625289; AAA52700.1; -
 CC EMBL: M64283; AAA52699.1; -
 CC EMBL: S76436; AAB21121.2; -
 CC EMBL: L22078; -; NOT_ANNOTATED_CDS.
 CC HSSP: P00740; IIXA.
 CC MIM: 142461; -
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR000742; EGF_2.
 CC InterPro: IPR001438; EGF_II.
 CC InterPro: IPR003006; Ig_MHC.
 CC InterPro: IPR003598; Ig_C2.
 CC InterPro: IPR002172; LDL_recept-A.
 CC InterPro: IPR000034; Laminin_B.
 CC InterPro: IPR002049; Laminin_G.
 CC InterPro: IPR001791; Laminin_G.
 CC InterPro: IPR000082; SEA.
 CC Pfam: PF00008; EGF_4.
 CC Pfam: PF00047; Ig_22.
 CC Pfam: PF00052; Laminin_B_3.
 CC Pfam: PF00053; Laminin_EGF_8.
 CC Pfam: PF00054; Laminin_G_3.
 CC Pfam: PF00057; LDL_recept_a; 4.
 CC Pfam: PF01390; SEA; 1.
 CC PRINTS: PR00010; EGFLOOD.
 CC PRODOM: PD003031; Laminin_B; 3.
 CC SMART: SM00180; EGF_Lam; 6.
 CC SMART: SM00001; EGF_Like; 8.
 CC SMART: SM00408; IGC2; 22.
 CC SMART: SM00281; Lamb; 3.
 CC SMART: SM00282; Lamb; 3.
 CC SMART: SM00192; LDLa; 4.
 CC SMART: SM00200; SEA; 1.
 CC PROSITE: PS00022; EGF_1; 9.
 CC PROSITE: PS01186; EGF_2; 5.
 CC PROSITE: PS01248; LAMININ_TYPE_EGF; 11.
 CC PROSITE: PS01209; LDLRA_1; 4.
 CC PROSITE: PS50068; LDLRA_2; 4.
 CC PROSITE: PS50024; SEA; 1.
 CC Signal: Basement membrane; Proteoglycan; Repeat; Glycoprotein;
 CC Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
 CC Extracellular matrix; EGF-like domain.
 CC SIGNAL 1 21 POTENTIAL.
 CC CHAIN 22 4393 BASEMENT MEMBRANE-SPECIFIC HEPARAN
 CC DOMAIN 80 194 SULFATE PROTEOGLYCAN CORE PROTEIN.
 CC FT 197 236 LDL-RECEPTOR CLASS A 1.

FT	DOMAIN	283	321	LDL-RECEPTOR CLASS A 2.
FT	DOMAIN	323	361	LDL-RECEPTOR CLASS A 3.
FT	DOMAIN	366	405	LDL-RECEPTOR CLASS A 4.
FT	DOMAIN	405	506	IG-LIKE C2-TYPE DOMAIN 1.
FT	DOMAIN	523	532	LAMININ EGF-LIKE 1 (N-TERMINAL).
FT	DOMAIN	533	732	LAMININ DOMAIN IV 1 (DOMAIN III A).
FT	DOMAIN	733	765	LAMININ EGF-LIKE 1 (C-TERMINAL).
FT	DOMAIN	766	815	LAMININ EGF-LIKE 2.
FT	DOMAIN	816	873	LAMININ EGF-LIKE 3.
FT	DOMAIN	881	925	LAMININ EGF-LIKE 4 (INCOMPLETE).
FT	DOMAIN	926	935	LAMININ EGF-LIKE 5 (N-TERMINAL).
FT	DOMAIN	936	1127	LAMININ DOMAIN IV 2 (DOMAIN III B).
FT	DOMAIN	1128	1160	LAMININ EGF-LIKE 5 (C-TERMINAL).
FT	DOMAIN	1161	1210	LAMININ EGF-LIKE 6.
FT	DOMAIN	1211	1267	LAMININ EGF-LIKE 7.
FT	DOMAIN	1277	1326	LAMININ EGF-LIKE 8.
FT	DOMAIN	1327	1336	LAMININ EGF-LIKE 9 (N-TERMINAL).
FT	DOMAIN	1337	1531	LAMININ DOMAIN IV 3 (DOMAIN III C).
FT	DOMAIN	1532	1564	LAMININ EGF-LIKE 9 (C-TERMINAL).
FT	DOMAIN	1565	1614	LAMININ EGF-LIKE 10.
FT	DOMAIN	1615	1672	LAMININ EGF-LIKE 11.
FT	DOMAIN	1679	1773	IG-LIKE C2-TYPE DOMAIN 2.
FT	DOMAIN	1774	1867	IG-LIKE C2-TYPE DOMAIN 3.
FT	DOMAIN	1868	1957	IG-LIKE C2-TYPE DOMAIN 4.
FT	DOMAIN	1958	2053	IG-LIKE C2-TYPE DOMAIN 5.
FT	DOMAIN	2054	2153	IG-LIKE C2-TYPE DOMAIN 6.
FT	DOMAIN	2154	2246	IG-LIKE C2-TYPE DOMAIN 7.
FT	DOMAIN	2247	2342	IG-LIKE C2-TYPE DOMAIN 8.
FT	DOMAIN	2343	2438	IG-LIKE C2-TYPE DOMAIN 9.
FT	DOMAIN	2439	2535	IG-LIKE C2-TYPE DOMAIN 10.
FT	DOMAIN	2536	2631	IG-LIKE C2-TYPE DOMAIN 11.
FT	DOMAIN	2632	2728	IG-LIKE C2-TYPE DOMAIN 12.
FT	DOMAIN	2729	2828	IG-LIKE C2-TYPE DOMAIN 13.
FT	DOMAIN	2829	2926	IG-LIKE C2-TYPE DOMAIN 14.
FT	DOMAIN	2927	3023	IG-LIKE C2-TYPE DOMAIN 15.
FT	DOMAIN	3024	3114	IG-LIKE C2-TYPE DOMAIN 16.
FT	DOMAIN	3115	3213	IG-LIKE C2-TYPE DOMAIN 17.
FT	DOMAIN	3214	3300	IG-LIKE C2-TYPE DOMAIN 18.
FT	DOMAIN	3301	3401	IG-LIKE C2-TYPE DOMAIN 19.
FT	DOMAIN	3402	3490	IG-LIKE C2-TYPE DOMAIN 20.
FT	DOMAIN	3491	3576	IG-LIKE C2-TYPE DOMAIN 21.
FT	DOMAIN	3577	3671	IG-LIKE C2-TYPE DOMAIN 22.
FT	DOMAIN	3701	3847	LAMININ G-LIKE 1 (GLOBULAR DOMAIN V A).
FT	DOMAIN	3848	3883	EGF-LIKE 1.
FT	DOMAIN	3886	3924	EGF-LIKE 2.
FT	DOMAIN	3966	4104	LAMININ G-LIKE 2 (GLOBULAR DOMAIN V B).
FT	DOMAIN	4106	4143	EGF-LIKE 3.
FT	DOMAIN	4145	4178	EGF-LIKE 4.
FT	DOMAIN	4243	4391	LAMININ G-LIKE 3 (GLOBULAR DOMAIN V C).
FT	SITE	65	67	HEPARAN SULFATE (POTENTIAL).
FT	SITE	71	73	HEPARAN SULFATE (POTENTIAL).
FT	SITE	76	78	HEPARAN SULFATE (POTENTIAL).
FT	SITE	4151	4153	MEDIATES MOTOR NEURON ATTACHMENT (POTENTIAL).
FT	SITE	4301	4303	MEDIATES MOTOR NEURON ATTACHMENT (POTENTIAL).
FT	DISULFID	199	212	BY SIMILARITY.
FT	DISULFID	206	225	BY SIMILARITY.
FT	DISULFID	219	234	BY SIMILARITY.
FT	DISULFID	285	297	BY SIMILARITY.
FT	DISULFID	292	310	BY SIMILARITY.
FT	DISULFID	304	319	BY SIMILARITY.
FT	DISULFID	325	337	BY SIMILARITY.
FT	DISULFID	332	350	BY SIMILARITY.
FT	DISULFID	344	359	BY SIMILARITY.
FT	DISULFID	368	381	BY SIMILARITY.
FT	DISULFID	375	394	BY SIMILARITY.

Query Match 3.7%: Score 102; DB 1; Length 4393;
 Best Local Similarity 24.1%: Pred. No. 51;
 Matches 81; Conservative 38; Mismatches 137; Indels 80; Gaps 13;
 OY 48 IIVDTNVLTLFLPSGNEALPLIYKNGDI-----VRFRLKI-----QVYKKE 92

DB 304 -REVAKV-TDDKSVK-----AAKKOV-----SLSAVVLTEVKKHAGLPTHS 346
 QY 388 LQEVPEEGDLIIIFODGATKTPDKLONTSLYDSKIWT-----TKNGGRKVA----- 435
 DB 347 LQDLHMDTDKEISSKOTFRTOFYITREPDVKEMVASYDKRTKKPSHHGACAKGCE 406
 QY 436 ----VHEVAKNGCILPLSN---ECLLIEG-----TLSEICKUS 467
 DB 407 NIFVOFLVKKASTOLNNNTYRVLLYTODGLGAFNFVAKPDMLYKNNDARKKLEYNELL 466
 QY 468 NKENS-----VIPVRSQ 479
 DB 467 TKFNSTYDAVVERRRNG 482

RESULT: 5

GAT1_YEAST STANDARD: PRT: 510 AA.

AC P43574: 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TRANSCRIPTIONAL REGULATORY PROTEIN GAT1.
 GN GAT1 OR YFL021W.
 OS Saccharomyces cerevisiae (baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CRP8;
 RX MEDLINE=96182087; PubMed=8622686;
 RA Coffman J.A., Rai R., Cunningham T., Svetlov V., Cooper T.G.;
 RT 'Gat1p, a GATA family protein whose production is sensitive to
 nitrogen catabolite repression, participates in transcriptional
 activation of nitrogen-catabolic genes in Saccharomyces cerevisiae.';
 RL Mol. Cell. Biol. 16:847-856(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=95400292; PubMed=7670463;
 RA Murkami Y., Naitou M., Hagihara H., Shidata T., Ozawa M.,
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
 RA Yamazaki M., Tashiro H., Eki T.;
 RT 'Analysis of the nucleotide sequence of chromosome VI from
 Saccharomyces cerevisiae.';
 RL Nat. Genet. 10:261-268(1995).
 CC -1- FUNCTION: POSITIVE REGULATOR OF MULTIPLE NITROGEN CATABOLIC GENES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS 1 GATA-TYPE ZINC FINGER.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: 027344; AAB03516.1;
 DR EMBL: D50617; BAA09217.1;
 DR HSSP: P17429; SGAT.
 DR SGD: S0001873; SGAT.
 DR InterPro: IPR000679; ZNF_GATA.
 DR Pfam: PF00320; GATA.1.
 DR PRINTS: PR00619; GATAZNFINGER.
 DR SMART: SM00401; ZNF_GATA.1.
 DR PROSITE: PS00344; GATA_ZN_FINGER_1; 1.
 DR PROSITE: PS00114; GATA_ZN_FINGER_2; 1.
 KM Transcription regulation: DNA-binding: Zinc-finger: Nuclear protein.
 FT ZN_FING 310 334 GATA-TYPE.
 FT DOMAIN 151 158 POLY-ASP.
 SO SEQUENCE 510 AA: 56327 MW: 62080542695F35F CRC64;

Query Match 4.5%; Score 122; DB 1; Length 510;
 Best Local Similarity 17.9%; Pred. No. 0.065;
 Matches 80; Conservative 54; Mismatches 117; Indels 196; Gaps 15;

QY 209 HIRLQNLITDILVYDNHV-HVARSKVGSEFLRIYSLHTKQSMNSEQNTLSLEFHLHG 267
 DB 27 YIHCTRCYVNTITMSTNRPNLDPDLNLNK--EIDWLYSSAOKILPDSNRILNLISWRLHN 84
 QY 268 GTSYGRGIRVLPESNDV----- 286
 DB 85 RTSFHRIRINIMQHSINDFSAPFASGVNNAAGPCNNLDITDINOFPPLSDMLNNS 144
 QY 287 -----QLKDLSEANLTANQHS-----DVIGQSEPDSEF 315
 DB 145 VEFNVFDDDDDDVDVETHSHVHSDLLNDMDSASQASHNAGFPNFLTSCSSSFDHFI 204
 QY 316 -----PSSGVSILYEVKCO---QLSATILTDHQYLEPTPLCA 350
 DB 205 FTNNLPFLNNNSINNHNSSHNNSPSIANNTNANTNTSASTNTNSPLLRNPS 264
 QY 351 ILKQAPQOYRIRAKLRSYKPRPLFQSVYLHCPCHLIQEVPEEGDLIIIFODGAT---- 406
 DB 265 IYKPGSRNRSSVKKKKPKALKIKSSSTV-----QSSATPPSN 301
 QY 407 --KTDPVKLONTSLYDSKIWTTRKQKRVAVHEVYKNGILPLNSEC--LLIEGCT--L 460
 DB 302 TSSNPDIKSNCTTSTPLM-RKDPK-----LPLCNACGLFLKHLVTRPL 347
 QY 461 S-----EICKSNKFNSTVIVR 477
 DB 348 SLKTDIIKRRQSSRTKINNNTTPPPSSSLNPGACGKKKNTYASVASKRKSLSNIVAPLK 407
 QY 478 SCHEDELIDLSP---FLIGDTIHH 500
 DB 408 S--QDIPPKIASPSIPQYLRSTNRHH 432

RESULT: 6

EX02_BPT4 STANDARD: PRT: 560 AA.

AC P04522: 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-JAN-1990 (Rel. 13, Last annotation update)
 DE EXONUCLEASE SUBUNIT 2 (EC 3.1.11.-) (PROTEIN GP46).
 GN 46.
 OS Bacteriophage T4.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
 OC T4-like phages.
 OX NCBI_TaxID=10665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85257446; PubMed=4018026;
 RA Gram H., Rueger W.;
 RT 'Genes 55, alpha 9, 47 and 46 of bacteriophage T4: the genomic
 organization as deduced by sequence analysis.';
 RL EMBO J. 4:257-264(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kuter E., Aritsaka F., Kunisawa T., Tsugita A., Mosig G.,
 RA Mesyanzhinov V., Ruger W., Sildham T., Thomas E.;
 RT 'Bacteriophage T4 genome analysis.';
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 478-560 FROM N.A.
 RA Hsu T., Karam J.;
 RL Submitted (APR-1987) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: EXONUCLEASE INVOLVED IN PHAGE DNA RECOMBINATION,
 CC REPLICATION, AND REPAIR.
 CC -1- SUBUNIT: CONSISTS OF TWO SUBUNITS, GP47 AND GP46.
 CC -1- SIMILARITY: STRONG TO T5 PROTEIN D13 AND TO YEAST RAD52.

Db 69 VNVFFSQNFEDLPIIQRWDIVRHARLQHYNDAKOLNVNRYSSMCLFIGNDKEAP 128
 Oy 114 IIPRT-----SKRYF-----NETTDDH--KNVEALRVASTHMSPTWL--LKLC 154
 Db 129 LERKVENEDOTNNYFTPTPNFSGKSTGTQGHGKILKDKKSKDYFSNNDVEQYKKA 188
 Oy 155 DVQPMOT---PULTCOLLGAKEVDCASFLKKYWDG-----RTPPSFVLIDOL 201
 Db 189 DIETAMKNTKIDFLAKKVEISDNDYTNVSLNDSTGTWIGHLKRFKPPH----- 240
 Oy 202 VLEGDSHRIHLONLITIDLVYDNHNVHVARSLKVSFLRILYSLHTKLOSNSEN 255
 Db 241 LVKGDVLRITSVAKEDNSLIFSISHNLK-----FFSSSIHKKLKSSISSD 288

RESULT 2
 TEER_EUPCR
 ID TEER_EUPCR STANDARD: PRT: 460 AA.
 AC 006183;
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DE TELOMERE-BINDING PROTEIN HOMOLOG.
 OS Euplotides crassus.
 OC Eukaryota: Alveolata; Ciliophora; hypotrichs; Euplotidae; Euplotidae;
 OC Euplotidae.
 OX NCBI_TaxID=5936;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93126105; PubMed-1480483;
 RA Wang M., Skopp R., Scofield M., Price C.;
 RT "Euplotides crassus has genes encoding telomere-binding proteins and
 telomere-binding protein homologs.";
 RL Nucleic Acids Res. 20:6621-6629(1992).
 CC -1- FUNCTION: MAY BIND TELOMERIC TAG4 SEQUENCES.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT
 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M96819; AAA29128.1; -
 DR PIR: S35525; S35525
 DR InterPro: IPR003415; Telo_bind_alpha.
 DR Pfam: PF02307; Telo_bind_alpha; 1.
 DR KJ: DNA-binding; Nuclear protein; Telomere; Multigene family.
 SQ SEQUENCE 460 AA: 53360 MW: EDIC141385A0B5FE CRC64;

Query Match 5.7%; Score 156.5; DB 1; Length 460;
 Best Local Similarity 21.4%; Pred. No. 0.00012;
 Matches 116; Conservative 74; Mismatches 196; Indels 155; Gaps 24;

Oy 9 YIYPLNQL--KGGIVNVGVKFFKPPYLSKGTDCSVVITVDOT-NVK-----LTC 59
 Db 19 YETIEISIEENASINFAVVIDACFPYKVDKRYKWLKVIDITHNKEGDNDFAIV 78
 Oy 60 LIFSAGTEALPIYKNDIYFRHLKIOYVKKETOGIT-----SSGASLIFETGLAPI 114
 Db 79 ALQSRKEDLPIDQCDIIRHRAEYN-YKDDHYFKLNKSYSSSMALTSADDEVAPV 137
 Oy 115 IP-----RTSKYFETEDHKVNEALRVMASTHMSPTWLKLCDDVQPMQYFD 163
 Db 138 IKDGDDEFTYRSYVSGKQYNFQDQDKLKNTRAMKNSYFAKN-----DV----- 183
 Oy 164 LTCOLLGAKEVDCASFLKKYWDGTRTPPSFVLIDOLVLEGDSHRIHLONLITDLYV 223

Db 184 -----IDEMTTPLSQARQEBGFENNVGKTYQIYHRDYTSDLRK 224
 Oy 224 DN-----HVVHVARSLKVGSLFRLY-SLHTKLOSNSENQTMLSLFFHLHGGTSGRGIRV 277
 Db 225 DISKATMFLYLR-----KFPRLYEGVILKIRSVINDS-----TERRCLEL 268
 Oy 278 LPESN---SDVDQLKDLSEANLTAHNSD-----VICQSPDDSPSSGSVSLYEYER 328
 Db 269 APMNSMTFTVPSRLAKSDQISLSPDKVDKELIKVLTPE----- 311
 Oy 329 COOLSATILTDHOYERTPLCAILKOKARQO---YRIRAKLRSPKRLFOYKLCPCKH 386
 Db 312 --VLATTTCGDISELPLTFLSEFEDYTDKDAVFARFSLIKITDRVEDYVEETPR-- 367
 Oy 387 LQGVPHGDDLDIIFQDQATKT-PDYKLO-----NTSLYDS--KIWT-TKQKGRKVAV 436
 Db 368 -----GAPRSKPYKYQFLKIDPSTALNDMLYKIKYISHDGLKKEFF- 409
 Oy 437 HEVKNNGILPLSNECLLIEGGLSEICKLSKFNKSVIPIRSGHEDLELDLSAPFLIOG 496
 Db 410 -----PGVDPSSAQ-----TPSGHSLK-RKYASTLKNFVHIDAVLEKGAFFIRD 455
 Oy 497 T 497
 Db 456 T 456

RESULT 3
 TEER_OXYNO
 ID TEER_OXYNO STANDARD: PRT: 495 AA.
 AC P29549;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE TELOMERE-BINDING PROTEIN ALPHA SUBUNIT (TELOMERE-BINDING PROTEIN 56
 KDA SUBUNIT) (TEER ALPHA).
 GN MAC-56A AND MAC-56K AND MAC-56S.
 OS Oxytricha nova.
 OC Eukaryota: Alveolata; Ciliophora; hypotrichs; Stichotrichida;
 OC Oxytrichidae; Oxytricha.
 OX NCBI_TaxID=5945;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE-92035001; PubMed-1840510;
 RA Gray J.T., Celandier D.W., Price C.M., Cech T.R.;
 RT "Cloning and expression of genes for the Oxytricha telomere-binding
 RT protein: specific subunit interactions in the telomeric complex.";
 RL Cell 67:807-814(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92249771; PubMed-1577273;
 RA Mitcham J.L., Lynn A.J., Prescott D.M.;
 RT "Analysis of a scrambled gene: the gene encoding
 RT alpha-telomere-binding protein in Oxytricha nova.";
 RL Genes Dev. 6:788-800(1992).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE-99091054; PubMed-9875850.
 RA Horvath M.P., Schweiker V.L., Bevilacqua J.M., Ruggles J.A.,
 RA Schultz S.C.;
 RT "Crystal structure of the Oxytricha nova telomere end binding protein
 RT complexed with single strand DNA.";
 RL Cell 95:963-974(1998).
 CC -1- FUNCTION: MAY FUNCTION AS PROTECTIVE CAPPING OF THE SINGLE-
 CC STRANDED TELOMERIC OVERHANG. MAY ALSO PARTICIPATE IN TELOMERE-
 CC LENGTH REGULATION DURING DNA REPLICATION. BIND SPECIFICALLY TO THE
 CC 14G4-CONTAINING EXTENSION ON THE 3' STRAND AND PROTECT THIS REGION
 CC OF THE TELOMERE FROM NUCLEASE DIGESTION AND CHEMICAL MODIFICATION.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- MISCELLANEOUS: THE SEQUENCE OF THE A (OR ALANINE) VERSION IS
 CC SHOWN. THE S (OR SERINE) VERSION DIFFERS IN ONLY TWO POSITIONS,
 CC AND THE K VERSION IN ONLY ONE. THERE MAY BE OTHER VERSIONS IN

OY 164 -LTCOL-LGKAEVDGASFLTKVMDGTRTPPSRWLVLO----DLVLEGD----LSHRL 213
DB 228 KFSCHIRGNPKRLRSTTVYKFAKEIP-----VIVENNSTVLYVERPTCLIKVNFKR 282
OY 214 ONLT--ID-ILVYDNH--VHVARSLAVG--SFLRYSLSHTKLOS--MENSENOMLSLEP 263
DB 283 ANITWTFDSFLHDEKEGIIYITNEERKKGKDFELKSVLIRVSHSNKPAOSDNLTIMCM-- 340
OY 264 HLHGISTYGRGIRVLPESNDVD-OLKOLESAN--LIANOHSVDYICQSPDPS-----F 315
DB 341 ----ALSPVPGKVMKNISSKRTIFLLGSEISSTOPLSVTE-STLDIQSPASSVSPARY 395
OY 316 PSSGSVSLYEV 326
DB 396 PATSVTLVDV 406

RESULT 12
T00342
hypothetical protein KIA0580 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00342
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: 214086; MUID:98290545
A:Accession: T00342
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1044 <NAC>
A:Cross-References: EMBL:AB011152; NID:93043683; PIDN:BAA25506.1; PID:93043684
A:Experimental source: brain; clone HJ0601
C:Genetics:
A:Note: KIA0580

Query Match 4.0%; Score 110.5; DB 2; Length 1044;
Best Local Similarity 20.2%; Pred. No. 3.4;
Matches 102; Conservative 68; Mismatches 171; Indels 143; Gaps 26;

OY 49 IYDQIVNKLTCLL--FSGNYEALPIIYKNGDIVRHRRLKIYKKKETOISSTGASLT 105
DB 255 LLEETN-KKMCVLEGGFLSTENDKSTTPNGTININEVICLAHKEDFYLTNPPIF-- 1 310
OY 106 FEGTLCAPLIIPRTSYFNTTEDHKVKEALRYMA--STHMSPTLTKLCDVQPMQYF 162
DB 311 FE-----IYLPSEVFLFGAETS--QOQRKMTETIAKHFPPLA-ENLTAD----Y 355
OY 163 DLTCQLGKAENDGASFLKVM--DGTPTPFSWRVLIODLVLEGDLSHHRLONLTD 219
DB 356 DLTGOLFETK-DCHALDQWRKGMFAMDKSSLSHF-----CLQMOEVOGDMRLRLRLTETIS 409
OY 220 ILVYDNH--VHVARSLAVGSLRIRYSLSHTKL-----QSMNSENOT 257
DB 410 TMQONEKELDVLVLENGRGLYIHG-HTKLDTVMTAIEKAAGTQGNALQDOOLSKNDV 468
OY 258 MSLFELHNGISTYGRGIRVLPESNSD-----VDLKKDLESANLTANOH--SDVIC- 307
DB 469 PIVNSCIAFVTOYGLGCKIYKNGDPLHISELSFKKARSFKLRACKHOLEDTAV 528
OY 308 ----QSEPDSPFSSGSVSLY-----EVERCOOLSATILLDHOYLETPICAIL 352
DB 529 LKSFSLDIDDALITK--ELYPWISALDQDOKERAKKYG-A-FIRSLPGVNRATIAAI 584
OY 353 KQAKPOQYRIARLASYKPRRLFOYVKLHCPKCHLQEVNHEG--DLDIDFQGGATKTPD 410
DB 585 E-----ELRYO-----KC-----SEINMNAHNLALV----- 608
OY 411 VKLQNTSLYDSKIMITKNQGRKVAVHFVKNNGILPLSNCELLILLEGTLSEICKLSNKF 470
DB 609 -----SSCLPOTKGTGOTSEEVNVEDLI 630

OY 471 NSVIPVRSGHED-LELIDLSAPFL 493
DB 631 NNYVEIFEVKEDQVQKQMDIENSFTI 654

RESULT 13
T21403
hypothetical protein F26D2.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T21403
R:McMurtry, A.
submitted to the EMBL Data Library, November 1996
A:Reference number: T21418
A:Accession: T21403
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1217 <MIL>
A:Cross-References: EMBL:281513; PIDN:CAB04180.1; GSPDB:GN00023; CESP:F26D2.10
A:Experimental source: clone F26D2
C:Genetics:
A:Gene: CESP:F26D2.10
A:Map position: 5
A:Intons: 29/1; 568/2; 608/3; 746/1; 871/3; 910/1; 985/1; 1018/3; 1124/1; 1170/3
C:Superfamily: Caenorhabditis elegans hypothetical protein F36D3.5

Query Match 4.0%; Score 108; DB 2; Length 1217;
Best Local Similarity 20.6%; Pred. No. 6.7; Mismatches 166; Indels 172; Gaps 30;

OY 45 SVYTIYDQIVNKLTCLL-----FSG-----NYEALPIIYKNGDIVRHRRLKIYKK 91
DB 483 AYSVLDQLOVNASNSIYKRLIRRYKGIYVNSYFELSTIKIARIAS-----ELQVQK 538
OY 92 ETQGITSSGFSALTFEGTLCAPLIIPRTS-----SKYFNTTEDHKVKEALRYASTH 143
DB 539 TAENITVS-----DGNILGDLISNLSLVETSKCFQYKTPPLESTVSKIK--SLR 588
OY 144 MSPSKTLKLCIDVQPMQYFDLTCLGKAEVDGASFLKVMQGT-----RTP 190
DB 589 MFPDGETISM--LEID--LLGKYOSNLTLVEKEKSTSHVNSTNNPVLLKTP 640
OY 151 FFSWRVLIODLVLEGDLSH--HR-----LQNLITD--ILVYDNH----- 228
DB 641 ---QALINKIGTSTVLE-DLENVRQHOFLYSIRNFTADYKVIYGENLLEIDPEKK 695
OY 229 -----VARSIAVGS--FLRIRYSLSHTKQSMNSENOTMLSLFHL--HGCT- 269
DB 696 LRELLEBADQNGFASGLRNGSLTEMARVFEAKIKIGIGDPRKSMANVSINLQASGTD 755
OY 270 --SYGRGIRVLPESNDVDOLKKDLESANLT-----ANOHSDVI--QSEPDPS 314
DB 756 QLAASVFKIRIELNDLPKAKHEASLSKASAVTYSLSIAYYDEIFGNSKSKILTEKIDNS 815
OY 315 FP-----SSGSV-----SLY-----EVERCOOLSATIL 337
DB 816 FPMIYILLICIGISILLIGFCVAVGYFTNGRAKYNNLYYFGKPLDERRRRYS--LF 873
OY 338 TDHOYLETPICA--ILKQAKPOQYRIARLASYKPRRLFOYVKLHCPKCHLQOE-- 390
DB 874 LQMDGKNTLLDVREINKNTNWKAAVKGSIYTAIK--FGTALMAATKSALPEIYKI 930
OY 391 -VPHEGDLDIIFQDGATKTPDVQLQNTSLYDSKIMITKNQGRKVAVHFVKNNG 443
DB 931 LIEHGADLTLL--NTLNKTPREQMIPIDSPDPSDKIEXLEKIKKL--YKNKANG 979

RESULT 14
T18410
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) II - malaria parasi
C:Species: Plasmodium falciparum

Db 345 ELKQTSVNEITAHVINEPASIKLTTITLHAPLQNLKPKRKHRLRVQVDFMPKSLQF 404
Oy 378 VKLHCPKCH-----LLOEVPHEGDDLIIF--QDGAITKTPDVLONTSLD 420
Matches 116; Conservative 74; Mismatches 196; Indels 155; Gaps 24;
Db 405 AVLSQPSSTVYWMFALLVROVSNV-TLPVIFPDSDAEELINSSKTOPCNLAD 455

RESULT 2
S35524
telomere-binding protein, 51k - Euplotes crassus
C:Species: Euplotes crassus
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-Dec-1999
C:Accession: S35524
R:Wang, W.; Skopp, R.; Scofield, M.; Price, C.
Nucleic Acids Res. 20, 6621-6629, 1992
A:Title: Euplotes crassus has genes encoding telomere-binding proteins and telomere-binding protein
A:Reference number: S35524; MUID:93126105
A:Accession: S35524
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-446 <MAN>
A:Cross-references: EMBL:M96819; NID:g159023; PID:g159024
C:Genetics:
A:Genetic code: SGC9
A:Introns: 9/3
C:Keywords: DNA binding; nucleus

Query Match 6.0%; Score 164; DB 2; Length 446;
Best Local Similarity 23.8%; Pred. No. 7.2e-05;
Matches 70; Conservative 55; Mismatches 107; Indels 62; Gaps 13;

Oy 8 NYITPLNQL--GGTIVNVYGVKFFKPPYLSKGTID-YCSVVTIVDQTNV-----K 56
Db 11 HGYSDLSIKKEGEEDQYHEGVVVIDASEFY--KGEKRYVTCTKVPDSSVAKGKLTNT 68
Oy 57 LTCLFSGNTEALPIIKNDIVAFHRLKIOVKKETGQITSSGFS--LTFEGTLGAP 113
Db 69 VNVFFSOFNEDLPITROVDIVVHRAQLQHTDAKQALNVNMYRSMSCLEFGNDKEAP 128
Oy 114 IIPRT-----SSKYF-----NFTEDH--KMYEALRVMASTMSPSMTL--LKL 154
Db 129 LEPEVENEDGNNFYSTPYNFSKGSFTQEGHETKILKDKMKSQDYFSNNDVVEGYKKA 188
Oy 155 DVQPMQY----FDLTQQLKAEVDGASFLKQVMDGT-----RTPPSMVLVLOD 201
Db 189 DIETAMKNTKIDFLAKVTEISDNDQYTNVSLNDSTGQTWTGHLFRKRPVH----- 240
Oy 202 VLBCDLSHHRQLMLTIDILVYDNHVVARSCLKVGSFLRIYSLHTKLSQSMSEN 255
Db 241 LVKGDVLRIRKSVSAKEDNSLIFSSHNLK-----FFSSSIHKKLKSSISSD 288

RESULT 3
S35525
telomere-binding protein - Euplotes crassus
C:Species: Euplotes crassus
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-Dec-1999
C:Accession: S35525
R:Wang, W.; Skopp, R.; Scofield, M.; Price, C.
Nucleic Acids Res. 20, 6621-6629, 1992
A:Title: Euplotes crassus has genes encoding telomere-binding proteins and telomere-binding protein
A:Reference number: S35524; MUID:93126105
A:Accession: S35525
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-460 <MAN>
A:Cross-references: EMBL:M96819; NID:g159025; PID:g159026
C:Genetics:
A:Genetic code: SGC9
A:Introns: 18/3; 377/3
C:Keywords: DNA binding; nucleus

Query Match 5.7%; Score 156.5; DB 2; Length 460;
Best Local Similarity 21.4%; Pred. No. 0.00029;
Matches 116; Conservative 74; Mismatches 196; Indels 155; Gaps 24;

Oy 9 YIYPLNQL--KGGTIVNVYGVKFFKPPYLSKGTIDYCSVVTIVDQTNV-----LTC 59
Db 19 YETIEGSIIEENENASINFVAVVIDACFPYKVDKMYCLKVIIDTTHNVKEDDNFALV 78
Oy 60 LTFSGNTEALPIIKNDIVAFHRLKIOVKKETGQIT-----SSGFASLTFEGTLGAP 114
Db 79 ALQSRKFEPLPIORCDIIRVHRAEYN-YKDOHYFKLMYSYSSMALFSADQEAPEV 137
Oy 115 IP-----RTSSKYFNTTEDHKAWEALRVMASTMSPSMTLKLCDVQPMQYED 163
Db 138 IMKGGDFTFRSYAYSQKYNFIDQDKLTKLRANKMSYFAFN-----DV----- 183
Oy 164 LTCQLGKAEVDGASFLKQVMDGTTRPPSPMRYLIDLVLEGDLSHHRQLNTIDILVY 223
Db 184 -----IIDENYTPLSQARQEGDFNVGVKVTQIVHRYDTTSLRLVK 224
Oy 224 DN-----HVVARSCLKVGSFLRIY-SLHTKLSQSMSENQTMLSLEPHLGGTSGYGRIV 277
Db 225 DTSKATVFLTVSR--KPRLEGVYIKIRSVNIDSE-----TERECLEL 268
Oy 278 LPESN--SDVDQLKKDLESANLTANQHSN-----VICQSEDDSPSSGSVLYEVER 328
Db 269 APHSNIMTFVFPFSLKSLDSQISLSPDKVDKELIKVILTEP----- 311
Oy 329 CQQLSATIILTDHOLETFLPCAILKOKAPOO--YRIAKLRSPKPRRLQSVKLCPCCH 386
Db 312 --VLAITTFGVDSELPTELSEIFEDVTDQVAFRAFSILKITTPTVEVEYVEYEPK-- 367
Oy 387 LLOEVPHEGDDLIIFQDGAITK--PDVKLQ-----NTSLDS--KIWT--TKNKGKRVAV 436
Db 368 -----GARRSKPVYKVOFLINDPSALNDNLYKIYLSHGDGKEFF-- 409
Oy 437 HFVKNNGILPLNECLLIGGTLSEICKLSKFNSTYIPRSHEDELLDLSAPPLIG 496
Db 410 -----PCVDPSAQQ-----TPSGHSL--RRYASTLMKFNVIDAVLEKGAFFIRD 455
Oy 497 T 497
Db 456 T 456

RESULT 4
A41221
macronuclear alpha protein (alanine version) - Oxytricha nova
C:Species: Oxytricha nova
C>Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 02-Mar-2001
C:Accession: A41221; B41221
R:Gray, J.T.; Celander, D.W.; Price, C.M.; Cech, T.R.
Cell 67, 807-814, 1991
A:Title: Cloning and expression of genes for the Oxytricha telomere-binding protein
A:Reference number: A41221; MUID:92035001
A:Accession: A41221
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-495 <GRA>
A:Cross-references: GB:M68931; NID:g159811; PID:g159812
A:Accession: B41221
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-310, 'S', 312-455, 'E', 457-495 <GR2>
A:Cross-references: GB:M68930; NID:g159813; PID:g159814
C:Genetics:
A:Genetic code: SGC5
C:Keywords: DNA binding; nucleus

Query Match 5.4%; Score 148.5; DB 2; Length 495;
Best Local Similarity 19.6%; Pred. No. 0.0013;

MEDLINE 97189344
 REFERENCE 2 (bases 1 to 320)
 AUTHORS Green, E.D.
 TITLE Human chromosome 7 STSs (1997)
 JOURNAL Unpublished (1997)
 COMMENT

GDB: GDB:1318142
 Contact: Eric D. Green
 Genome Technology Branch
 National Human Genome Research Institute/NIH
 49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892
 Tel: 3014020201
 Fax: 3014024735
 Email: egreen@nhgri.nih.gov
 Primer A: CAGAAAAGTGTGATGAG
 Primer B: AGCTAAAGAGTGTGAG
 STS size: 96
 PCR Profile:

Presoak: 0 degrees C for 0.00 minute(s)
 Denaturation: 92 degrees C for 0.17 minute(s)
 Annealing: 55 degrees C for 1.00 minute(s)
 Polymerization: 72 degrees C for 1.00 minute(s)
 PCR Cycles: 35
 Thermal Cycler: PerkinElmer 9600

Protocol:
 Template: 30-100 ng
 Primer: each 1 uM
 dNTPs: each 200 uM
 Tag Polymerase: 0.05 units/uL
 Total Vol: 10 uL

Buffer:
 MgCl2: 1.5 mM
 KCl: 100 mM
 Tris-HCl: 10 mM
 NH4Cl: 5 mM
 pH: 8.6

The sequence for this STS was derived from a single sequencing read. For additional information about the NHGRI chromosome 7 mapping project, see <http://www.nhgri.nih.gov/DIR/GB/CHR7>. Also see Genomics 11:548-64 (1991) [MUID=92128937].

FEATURES
 source 1..320
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="7"
 /clone_lib="Eric D. Green"
 STS 181..276
 primer_bind 181..198
 primer_bind complement(259..276)
 BASE COUNT 103 a 47 c 52 g 112 t 6 others
 ORIGIN

Query Match 5.8%; Score 104.8; DB 11; Length 320;
 Best Local Similarity 92.4%; Pred. No. 8, 1e-15;
 Matches 109; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Oy 1588 catgattctgcaaatcttcacagatccagatcagaagtcgatgatgacct 1647
 ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 120 CATTTTAAAGACAAATCTTCACATTCACACATCAGAAAGTNCATGATGATGACCT 179
 Oy 1648 tcagaaagtgatgatgatgatgatgatgtttgtctccagagataaattgat 1705
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 180 TCAGAAAAGTGTGATGATGATGATGATGATGTTTGTCTCCAGNATAAAATGTGT 237

RESULT 10
 MPULM01/c 327650 bp DNA BCT 14-May-2001
 LOCUS MPULM01 327650 bp DNA BCT 14-May-2001
 DEFINITION Mycoplasma pulmonis (strain UAB CTIP) complete genome; segment 1/3.
 ACCESSION AL445563 AL445565

VERSION AL445563.1 GI:14089413
 KEYWORDS
 SOURCE Mycoplasma pulmonis.
 ORGANISM Mycoplasma pulmonis.
 Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;

REFERENCE
 AUTHORS Mycoplasmataceae; Mycoplasma.
 1 (bases 1 to 327650)
 Chamblaud, I., Hellig, R., Ferris, S., Barbe, V., Samson, D.,
 Gailson, F., Moszer, I., Dybvik, K., Wroblewski, H., Viar, A.,
 Rocha, E.P.C. and Blanchard, A.
 The complete genome sequence of the murine respiratory pathogen
 Mycoplasma pulmonis
 Nucleic Acids Res. 29 (10), 2145-2153 (2001)

TITLE
 JOURNAL Mycoplasma pulmonis
 MEDLINE Nucleic Acids Res. 29 (10), 2145-2153 (2001)
 REFERENCE 2 (bases 1 to 327650)
 AUTHORS Blanchard, A.
 TITLE Direct Submission
 JOURNAL Submitted (16-OCT-2000) Blanchard A., INRA, Centre de Recherche de
 Bordeaux, Institut de Biologie Vegetale Moleculaire, 71 avenue
 Edouard Bourleaux, BP 81, 33883, Villenave d'Ornon, Cedex, France

FEATURES
 source 1..327650

gene
 CDS

/organism="Mycoplasma pulmonis"
 /strain="UAB CTIP"
 /db_xref="taxon:2107"
 222..1607
 /gene="MYP0_0010"
 222..1607
 /gene="MYP0_0010"
 /standard_name="dnaa"
 /note="Identified by sequence similarity; possible:
 similar to SWISS-PROT:P34028 (DNAA_SPLIC) Blastp2 P=6e-47
 C=498"

gene
 CDS

/codon_start=1
 /transl_table=1
 /evidence=not_experimental
 /product="CHROMOSOMAL REPLICATION INITIATOR PROTEIN DNAA"
 /protein_id="CAC13174.1"
 /db_xref="GI:14089414"
 /translation="MSKSLDKIDLAVNTKLDIFESKINDSKLYONIKFELEITY
 TFDVKAIPADVDYIHLNLSDFKDKICIFOLEGNERETFTSGKROSSSPKSE
 NIOLVEESNNDELKKEFDNFLOSKEENAPLIGKEIIEPGLYVPLISGSGIG
 KTHLNAIGNEFLKPKPSIVKFLPSDFYKIPILSSNNINENAKFELEADLY
 MEDFOIFSIGNKRALINEIFEILDKRIMNKVITLSDDKLFMAFLFPORLYRLS
 SGLSVEIDPDKEDMKILKROLKIKNSPEKMKDLDPEVRNRSKSTFEELADLK
 VHFOLKROINEFNKRTISNFKNVEITKEKTEPTERTIEHANYKPKHLEIKSK
 SNLVYARDAIYLIREDLNVFSRIGLIFSNRDSTLFSNRKISOSINSRSTKTYI
 TLRSKINKLS"
 1741..2862
 /gene="MYP0_0020"
 1741..2862
 /gene="MYP0_0020"
 /standard_name="dnan"
 /EC_number="2.7.7.7"
 /note="Identified by sequence similarity; putative:
 similar to SWISS-PROT:P34029 (DP3B_SPLIC) Blastp2 P=3e-14
 C=438"

gene
 CDS

/codon_start=1
 /transl_table=4
 /evidence=not_experimental
 /product="DNA POLYMERASE III, BETA CHAIN"
 /protein_id="CAC13175.1"
 /db_xref="GI:14089415"
 /translation="MKLFLEKSRFLFLENTTKIIESNNSNYELRGAVFOYRDKIIL
 ISSNEDISKNEYLLDGIKNKTEGSDPLVNSLKNIKKCSITLLEKKNLKF
 SDLOSLEYNLNLDHERFPDKIFGLNSOQLTTEOFKOLKAVFRAADLSNQEILNS
 VNLAKNNILTLVATNARAIAMQIKTEDSOENLITNSKYKELISLSMSVNTLILSP
 GIFELKIKSGNLEIKTVYIEIPIYNVENVFPNKFNFYIHDKKELSLIDKVSIVNE
 KNGKILIEYNSKKEKKLKSYPDGFSEVSDNFVESEILKFEFINANYLAKES
 INVFDMSISIFTEKDRMVISSEFNLNKKLIALLNGH"
 2864..3085
 /gene="MYP0_0030"
 2864..3085

Db 869 TTTCGATCTTTGAGCTTTGAGGAGACTTTGGAGCCCTATCATACCTCGCACTTCAAGC 928

Qy 361 aagatatttaactcaactcaactgagaccacaatagttagaagcgtgttttgagca 420

Db 929 AAGTATTTTAACCTACTACTGAGGACCAACAAATGTAAGAACTTACTGTTTGGCA 988

Qy 421 tctaactcaatgtcacccgtcttgagacataactaaatgtgtgagtgctcaagccaatgca 480

Db 989 TCTACTCATATGTGACCGCTTGCGACATTACTAAATGTGTGATGTGACCAATGACAG 1048

Qy 481 tatttgacctgaactgttgaactccttgaggcaagaagtgagcagacatcttct 540

Db 1049 TATTGTGACCTGACTTGCTGACCTCTTGCGCAAGAGCAAGATGAGATGACCATCATTTCT 1108

Qy 541 cttaagatagtggatgagccagagacacatttccaacttgagagcttcaataacaagac 600

Db 1109 CTAAAGGATGGATGGACGACAGACACCATTTCCATCTTGAGAGCTTTAATACAGAC 1168

Qy 601 ctgtgtcttgaagtgatatttaagtcacatccacgagcttcaaaaacttgacataagacat 660

Db 1169 CTGTGCTTGAAGGTATTTAAGTCACATCCATCGGCTACAAATCTGACAAATAGACATT 1228

Qy 661 ttagtctagataaccatgttcaactgtgagcaagatcctctgaagtgtagaagcttctaga 720

Db 1229 TTAGCTAGAGATTAACCATGTTCTATGTGGCAAGATCTCTGAGGTTGGAACTTTCTTACA 1288

Qy 721 atctatagccttcaacaaacttcaactcaatgaatcagagaaatcagaacatgttaagt 780

Db 1289 ATCTAATAGCCTTCAATCCAAACTTCAATCATGATCATGACAGATGACAAATGTTAAGT 1348

Qy 781 ttagagtttcaatcttcaactgagagtlaccagtttagcgtcggagaaatcagggtcttgccaaga 840

Db 1349 TTAGATTTTCAATCTTATGAGAGGTACCAAGTTATGTGGGGAATTCAGGCTCTTGCCAGAA 1408

Qy 841 agtaactcttgatgtgagatcaactgtaaaagatttagaactctgcaaatctgacaagccaat 900

Db 1409 ACTAATCTTGATGTGATGATCACTAGTGAAGAAAGATTTAGAATCTGCCAATTTGACAGCTTAT 1468

Qy 901 cagcatctcagatgttactctgttcaactagaacctgacagacagcttctcaagctctgagta 960

Db 1469 CACCATCTTCAGATGTATCTGTCAATGAGAACCTGATGACAGCTTTCCAAAGCTCTGATGATA 1528

Qy 961 gtaactcttcaactgagagtagaagaatgttcaacagctatctgttcaactaactcaagta 1020

Db 1529 GATATCTTGTATGAGGTATGAAGAGATTCACACCTATCTGTCTACATTTACAGATTCAT 1588

Qy 1021 cagatatttgagagagacacacactatgtgcatttctgaaacaaagaagctctcaacatac 1080

Db 1589 CATATATTGGAGAGGACACACACTGTGTGCATTTTGAAACAAAGACTCTCTCAACATAT 1648

Qy 1081 cagatccgagcaaaatgtgagtcataactaagccagagaagactatctagtcgtttaaact 1140

Db 1649 CGCATCCGAGCAAAATTTAGGTCATTAAGCCCGAAGACTATTTACAGCTGTGTAACCT 1708

Qy 1141 cactgcccataatgtcattctgtcgaagaagttccacatgaaggagcatttgatataact 1200

Db 1709 CATTGCCCTAAATGTCATTTACTGCAAGAAATTCACGTCGAGGCTATTTGGATATATTT 1768

Qy 1201 tctcagagatgtgcaactaaacccagatgttcaagctcaaaatacatcatatagat 1260

Db 1769 TTGACAGATGTGTCAAAATCCAGATGTCAAACTTCAAAATACACATTTATTCAT 1828

Qy 1261 tcaaaaacttgagcacacttaaaaaactcaaaaagacggaagatgacgttcaatttctgtgaaa 1320

Db 1829 TCAAAAATCTGACACACTAAATATCAAAAGAGCAAGATGACCAATTTATTTTGGA 1888

Qy 1321 aataatgtagtctccgcttcaaatgtagtctactcttgatagaagaggtacacac 1380

Db 1889 AATTAATGATTTCTCCCACTTTTCAATATGATGTCTACTTTTATAGAACAGAGTACACTC 1948

Qy 1381 agtgaataattgcaaacctctgacaagaatttaagtgtaactctcgttgagactgtgcccac 1440

Db 1949 AGTGAATTTGCAAACTCTGAAACAGTTTAATAGTAAATCTCTGTGAGATCTGGCCAC 2008

Qy 1441 gaagacctggaactcttgagacttccagcaccattcttatacaagagaacatacatcac 1500

Db 2009 GAAGACCTGGAACCTTTTGATCTTTTACGACCATCTTTTATACAGAAACATATCATCAC 2068

Qy 1501 tatggc----- 1506

Db 2069 TATGATGATGAACAGTGTCTAGTTTGAGATCATACAAATCTAATTTGTTGGTGAT 2128

Qy 1507 -----acgggtatctgtaccctcccaatat 1531

Db 2129 AAACATCTGCGATTCCTTCTTCTGTGCGAGAAAGTATGGATTTATGACCTCCAAATAT 2188

Qy 1532 gtgtgtgtatgaccttaccacttgatgatgagaagagtlactagaagccatctcatg 1591

Db 2189 GTGTTTGTATGACGTTTACACTTGATGATGGAACAGAGTACTAGAACCTATCTCATG 2248

Qy 1592 gatctgacaatcttctccagatctccagatcagaagaagttcgtatgatgatgacctcag 1651

Db 2249 GATTCGACAAATTTCTTCAGATCCACATCCAGATCGAAGTTCAATGAGATGACCTTCAG 2308

Qy 1652 aaaaagtgtgataatgataatgtgtgtgtctccaggaataaataatgatgacatat 1711

Db 2309 AAACGATGATGATATGATATGATATGATATGTTTGTCTCCAGGATTAATAATGATGATAT 2368

Qy 1712 ccgtggttggaatgcttcaatcaatgatacaaatgtcacaatgagacagataatcaat 1771

Db 2369 CCGTGTGTTGGAATGCTTCAATCAATGATCATGATGATGATGATGATGATGATGATGAT 2428

Qy 1772 tgcctcagatcttggacacacagcttgagaagatgtcaatctaa 1816

Db 2429 TCTATCAAGATTTTGGACACCAAGTTCAGAAATGATATCTAA 2473

RESULT 6

LOCUS AK026234

DEFINITION Homo sapiens cDNA: FLJ22581 fis, clone HS102573.

ACCESSION AK026234

VERSION AK026234.1 GI:10439019

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens human small intestine cDNA to mRNA, clone_Jib:HS1 clone:HS102573.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1752)

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

TITLE Watanabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H., Nakamura, T., Isogai, T., Ota, T., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.

JOURNAL NEDO human cDNA sequencing project

AUTHORS NEDO human cDNA sequencing project

TITLE Direct Submission

JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'-6' 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES

source 1..1752

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

SOURCE Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
clone.lib:NT2RM2 clone:NT2RM2001544.
ORGANISM Homo sapiens
REFERENCE Bukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
AUTHORS I (sites)
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shitatori,A., Sudo,H.,
Magatsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahara,K., Masuno,Y., Niinomiya,K. and Iwayanagi,T.
TITLE MEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2631)
AUTHORS Isogai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
Isogai, Heli Research Institute, Genomics Laboratory: 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT MEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing. Research Association for Biotechnology: cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Heli Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
source
1..2631
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone.lib="NT2RM2001544"
/clone.lib="NT2RM2"
/note="cloning vector: pME18SFL3; mRNA from uninduced NT2
neutonal precursor cells."
24..1928
CDS
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:7022355"
/protein_id="BAA91568.1"
/translation="MSLVPATNYYTPLNOLKGTIYNYGVYKFFKPYLSKTDVY
SVTVVDQTNVTKLCLEPSCNTEALPITIKNGDLYVRHRLIOYTKETOGITSSGFA
SLTEFTGLAPLIPRTSSKYFNTTEDHKWVLEALRVASHTHMSWTLLKCDVQPMQ
YFDLTCOLLKAEVDGASFLKVDGRTIPSPNRVYIODLYEGDLSHIRLONLTI
DIYVDNHNHVARSLKVGSLFRLISLHTRKQSMNSEQIMLSLEFHLGSGYSGRIR
VLPSNSVDQLKRDLESANLTANOHSDVICOSSPDSFSSGVSLEYVERCOOLSA
TILTDHOYLEPTPLCALIKOKAPOOYRIRATLRSYRPRRLOSUKLCPKCOLLEVP
HEGDLDIIFDGGATRTTPYKLONTSLYDSKWTTRKNOKRVANVHFNKNSILPLSNE
CLLIEGGTISEIKLSKNFNSVLPVRSGBEDLDELDSAPFLQGTIHHGCKOCSS
LRSIONLSLVDKTSWIPSSVFAALGIYVPMFTLDDGVEALVLMDSDFE
QIPASEVYLMDDDLDRSYDMIMDMFCPPKIDAYPMLECFIKSYNVTNGIDNOICYOI
FDTVAEVDYI"

BASE COUNT 826 a 473 c 491 g 841 t
ORIGIN

Query Match 94.4%; Score 1713.8; DB 9; Length 2631;
Best Local Similarity 95.2%; Pred. No. 0;
Matches 1814; Conservative 0; Mismatches 2; Indels 89; Gaps 1;

OY 1 atgtcttggttcagcaacaatatatatataccccctgaatcaacttaagtggt 60
|||||
DB 24 ATGCTTGGTTCAGCAACAATAATATATATACACCCCTGAATCAACTTAAGGTGT 83
OY 61 acaattgcaatgtctatgtgttgaaagctttaagcccccatatcaagaaga 120
|||||
DB 84 ACAATGTCAATGTCTATGTGTGTGAAGTCTTTAAAGCCCATATCAAGCAAGCA 143
OY 121 actgattatgtctcagtgtaactatgtgacccagacaatgtaaaacttaactgctg 180

|||||
DB 144 ACTGATTAATGTCTAGTGTGAATATGTGTGACGACGACCAATGAATGAATCTTGCTTG 203
OY 181 ctcttggttgaaactatgaagccctccaaatattataaaatggaatgtttcgc 240
|||||
DB 204 CTCCTTGTGGAAGCAATGAAGCCCTTCCAAATTTATATAAATGGAGATTTGTCGC 263
OY 241 ttcacagctgtaagaattcaaglatataaaaggagacacaggtatcaccaactctgc 300
|||||
DB 264 TTTCACAGGCTGAAGATTCAAGTATATAAAGAGACACTCAGGATATCACAGCTGCGC 323
OY 301 ttgcactcttgaaagcttgagggaactcttgaggccccctatcaactctgcacttaagc 360
|||||
DB 324 TTTCATCTTTGAGCTTTGAGGGAATCTTGGACCCCTTATCACTCTGCACCTTCAAGC 383
OY 361 aagattcttaactcactactgaagaccacaanaatgtgagaagccttaagctgttgagca 420
|||||
DB 384 AAGTATTTTAACCTTCACTACTGAGACGACCAAAATGTGATAGAGCTTACGTTTGGCA 443
OY 421 tctacatcatgtcacccgctcttgacacttaactaaatgtgtgattcagcaatgag 480
|||||
DB 444 TCTACTATATGTCCCGCTCTTGACATTAATAATTTGTGTGATGTTTCAGCCAAATGAC 503
OY 481 tatttgacctgtctgtcagctcttgaggcaagcagaagtgtgacgagacatcttct 540
|||||
DB 504 TATTTTACCTGACTGTGACACTCTTGGCAAGCAGAGAGGACATCAATTTCTT 563
OY 541 cttaagatgatgtgacacagacacacatctccatcttgagagagcttaataaagac 600
|||||
DB 564 CTAAAGGATATGATGAGCAGCAGACACACATTTCCATCTTGAGAGTCTTAATCAAGAC 623
OY 601 ctgttcttgaaagtgatgttaagtacacatcacatcagcgtcacaaatctgcaaatgacatc 660
|||||
DB 624 CTGTGTTGAAAGGATTTTATAGTCACATCCATGCGCTACAAATCTGCACATAGACATT 683
OY 661 ttaagctacgaataacatgtctcatgtgtgcaagatctctgaaagcttggaagcttctctaga 720
|||||
DB 684 TTATGTTATGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 743
OY 721 atctatagctctcatcacaactcacaatcaatgaattcagaagaatcagacaatgttaagt 780
|||||
DB 744 ATCTATAGCTCTCATACCAAACTTCAATCAATGAATTCAGAGATTCAGCAATGTTAAGT 803
OY 781 ttgaagttcatctctcatgtgaggtgacaggttagagtgagggaatcagaggtcttcagaa 840
|||||
DB 804 TTAGAGTTTCACTTTCATGAGAGTACAGATTACGCTCGGGAAATCAGGCTTTCGCCGAA 863
OY 841 agtaactctgtatgtgatacaactgaaagaattagaaatctgcaaatcttgacagccaat 900
|||||
DB 864 AGTAACCTGTGATGATGATCACTGAAAGAGATTTAGAAATCTGCAAAATTTGACAGCCAAAT 923
OY 901 cagcatctgaatgtatctatgtcatctcagaactctgacgacagcttccaagctctgagta 960
|||||
DB 924 CAGCATTTGAGTATGTTATCTGTCAATCAAACTTCAAGCAGACCTTCCAAAGCTCTGGATCA 983
OY 961 gtatcattatcagagtgtaagaagaatgtcaacagctatctgttaactatcatcagatcat 1020
|||||
DB 984 GTATCATTTATAGAGATGAAGATGATCAACAGCATCTGCTACAAATCTTACAGATCAT 1043
OY 1021 cagtaattggaaggaacacacatgtgtgcaatttgaaacaaaagctctcaacaaatc 1080
|||||
DB 1044 CAGTATTTGGAGAGACACCACTATGTGCCATTTTGAAGCAAAAAGCCCTCAACCAATATAC 1103
OY 1081 cgcacgcagacaaaattgaagtcataatagcccaagaagaactattcaagctgtttaaactt 1140
|||||
DB 1104 CGCATTCGACGCAAAATTTAGGTCATATAGCCCGACAGAACTATTTCAGTCTGTTAACTT 1163
OY 1141 catgcccataatgtcatctgtctgcaagaagttccacatgagggcgatlttgataaatt 1200
|||||
DB 1164 CATTCGCCATTAATGTCATTTGCTGCAAGAAGTTCACATGAGGGGATTTGGATATAATT 1223
OY 1201 ttccagatgtgtgcaactaaacccagatgtcgaagtacaaaatacactatattatgat 1260
|||||

Email: c9apbs-r@mail.nih.gov
 Tissue Procurement: ATCC/DCTD/DTF
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The J.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc-mgc@nhgri.nih.gov
 Shenchenko, Y., Weltherby, K.D., Beckstrom-Sternberg, S.M.,
 Benjamin, B., Blakesley, R.M., Bouffard, G.G., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
 Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Snyder, B., Stanlipop, S., Thomas, P.J.,
 Tjongson, E.E., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A.,
 Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the J.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 15 Row: 9 Column: 18
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 10434050.

FEATURES

Source

Location/Qualifiers

1..2029

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="MGC:10280 IMAGE:3955573"

/tissue_type="Skin, melanotic melanoma."

/clone_lib="NH_MGC_20"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

67..1971

/codon_start=1

/product="Unknown (protein for MGC:10280)"

/protein_id="AAH02923.1"

/db_xref="GI:12804139"

/translation="MSLVPAATNITPLNOLKGTIVVGVKFFKPPYLSKGTDC

SVVITVDQTNKLCILFSGNEALPIIYKNGDIYRFLKIOVKKETOGITSSGFA

YDLFTGLGAPIIPTSSKYNFTEDHKMVALMVAISHSPMTLLKLDVOPMO

YDLTQCLGAEVAGAFILKVMGTRPSPVNLIDVLLEDLSHILHOLTI

DLVYDNHVAHRSALKVGSFLIYSLHTRLOSNGENOTMLSEFLHSGYSYGRIR

VLPESNDVDOLKDLSEANLTAHNSHOVICOSEPDSPSSGSVSLYEVECCOOLSA

TILDYOLERPLCALITKORAKPOOYRIKAKRSKRPRLLOSVALHKKHCLDEVP

HEGDIIIFQBGATTPPYKIONISLIDSKIWITKNQKVAHFKNNGLPLPNS

CLLIEGGLSEICKLSNKFNSVIVPVRSGHEDLELDLSAPFLIGTIIHNGCKCSS

LSKIONLSLVKTSWIPSSVAEALGIYPLQVFMFTLDDGTGLVLAAYLMDSKRF

QIPASEVLMDDDLQKSDVMIMDMCEPCGICIKIDAYPMLECFIKSYNVINGTNOIQYI

FDTYAEVYI"

BASE COUNT

655 a 391 c 386 g 597 t

ORIGIN

Query Match

94.5%; Score 1717; DB 9; Length 2029;

Best Local Similarity

95.3%; Pred. No. 0;

Matches 1816; Conservative

0; Mismatches 0; Indels 89; Gaps 1;

0Y 1 atgtcttgggtccagcaacaatatataatacaacccctgaataactgaagtgat 60
 |||||||
 Db 67 atgtcttgggtccagcaacaatatataatacaacccctgaataactgaagtgat 126
 |||||||
 0Y 61 acgaattgcaatgctatgctgtgtgaagctcttaagcccccatactaaagaaga 120
 |||||||
 Db 127 ACAATTGTCATAGTCTATAGTGTGTGAAGTTCTTTAAGCCCCCATATATCAGCAAGA 186
 |||||||
 0Y 121 actgattatgtcagctgtgaatactgtgacccagacaagaatgaactgaagctg 180
 |||||||
 Db 187 ACTGATTATGTGCTACGTGTGAACATGTGTGACCAAGTGTAAACTACTTCCG 246
 |||||||
 0Y 181 cctcttaagtggaactaagaagcccttccaataattataaagaatggaatattgtgc 240
 |||||||
 Db 247 CTCTTTAGTGAAGAACTATAGAGCCCTTCCAAATATATTAATAAAGTGAATATGTGCC 306
 |||||||
 0Y 241 ttccacagctggaagattcaagatatataaagaagagcctcaggtatcaaccagctgc 300
 |||||||

Db 307 ttccacagctggaagattcaagatatataaagaagagcctcaggtatcaaccagctgc 366
 |||||||
 0Y 301 ttgcaccttgaagctgtgaaggaacttggagccctatacaactgcgaactgaagc 360
 |||||||
 Db 367 ttgcaccttgaagctgtgaaggaacttggagccctatacaactgcgaactgaagc 426
 |||||||
 0Y 361 aagattttaaactcaactcagcagcagcagcagcagcagcagcagcagcagcagcagc 420
 |||||||
 Db 427 AAGTATTTAACTTCTACTGACGACCAACCAAAATGTAAGACCTTACGTTGGGCA 486
 |||||||
 0Y 421 tctactcataatgcaacccgtcttggaactataaattgtgtgattcgaacccaatgag 480
 |||||||
 Db 487 TCTACTCATATGTCAACCGCTTGACATTAAGTAAATGTGTGATGTTGACCAATGAG 546
 |||||||
 0Y 481 tatttgaacctgaactgtgaactcttggaagcagcagcagcagcagcagcagcagcagc 540
 |||||||
 Db 547 TATTTGACCTGACTGTTGAGCTTTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606
 |||||||
 0Y 541 ctaaaagtatgagatgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 600
 |||||||
 Db 607 CTAAGAGTATGGATGGACACGACGACACATTTCCATCTTGAGAGTCTTAATACAGAC 666
 |||||||
 0Y 601 ctgtctctgaagtgatttaagtcacalcacgcgctacaaatctgacaatagacatt 660
 |||||||
 Db 667 CTGTCTCTGAAGTGATTTAAGTACATTCATCGGCTACAAATCTGCATATAGACAT 726
 |||||||
 0Y 661 ttagtctgaataaaccatttcatgtgtgaagatctctgaagcttggagagcttcttga 720
 |||||||
 Db 727 TTAGTCTAGCATAACCATTTGATGTGGCAAGATCTTGAGGTGGAAGCTTTTGTGA 786
 |||||||
 0Y 721 atctatagccttcatcaccac 780
 |||||||
 Db 787 ATCTATAGCTTCATATACCAACTTCAATCAATGAATTCAGAAATCAGACAAATGTAAG 846
 |||||||
 0Y 781 ttaagattcatcttcatatgagaggtacagctacggtcggggaatcaaggcttgcagaa 840
 |||||||
 Db 847 TTAGAGTTTTCATCTTATGAGAGTACCACTTACGCTGGGGGATAGGCTTTGGCAAGA 906
 |||||||
 0Y 841 agtaactctgaatgagtaactgaataaagaattagaactctgcaaatctgaacagcaat 900
 |||||||
 Db 907 AGTAATCTGATGTGATCACTGAAAAAGATTTAGAACTTGCAAATTTGACAGCCAT 966
 |||||||
 0Y 901 cagcaatcagatgtatctgtcaatcagaacctgacagacagcttccaagctctgagaca 960
 |||||||
 Db 967 CAGCAATTCAGATGTATCTGTCAATCAGAACCTGACAGAGCTTTCCAAAGCTTGATCA 1026
 |||||||
 0Y 961 gtaactatlaagaggaagaatgagcagcagcagcagcagcagcagcagcagcagcagcag 1020
 |||||||
 Db 1027 GTATCATTTATCGAGGTAGCAAAAGATGTCAACAGCTATCTGTACAAATCTTAACAGAT 1086
 |||||||
 0Y 1021 cagatttgaagagagac 1080
 |||||||
 Db 1087 CAGTATTTGAGAGGACACACATATGTGCTATTTGAAACCAAAAGCTTCCAAACATATAC 1146
 |||||||
 0Y 1081 cgcattcagagcaaatgtgaggtcatataagcccgagaagcatttcagctcgttaactt 1140
 |||||||
 Db 1147 GCGATTCGAGCAAAATGTAGGTCATATAAGCCAGAGACTATTTTCACGTGTAAACCT 1206
 |||||||
 0Y 1141 catggcccaaatgtcattgtctgcaagaagttccaatgagagcagcttggataaatt 1200
 |||||||
 Db 1207 CATGGCCCAATATGTCATTTGCTGCAAGAAGTTCCATGAGAGGGGATTTGGATTAAT 1266
 |||||||
 0Y 1201 ttcaagatgtgtcaactaaacccacagatgtcaagcttacaataacatcatatagat 1260
 |||||||
 Db 1267 TTTAGAGATGTGTCAACTAAACCCAGATGTCAAGCTCAAAATATCATTTATATATAT 1326
 |||||||
 0Y 1261 tcaaaatctgagacacataaataaagaagagcaagaagtagcggtcatttggtaa 1320
 |||||||
 Db 1327 TCAAAAATCTGACCACTAAATATCAAAAAGAGCAAGAGTGAAGTTCATTTTGTGAAA 1386
 |||||||
 0Y 1321 aataatgtatctccgcgttccaatgaaatgtcacttltgatagaagagatcacctc 1380
 |||||||

```

Db 663 www.....c..... 722
Oy 1722 aatgctcatcaagtcatacaatgcagaatcaatttcatacga 1781
Db 723 www..... 782
Oy 1782 ttlttgacaccac 1794
Db 783 wwwggcccccgc 795

```

RESULT 15

AAF58257 standard; DNA; 936 BP.

AC AAF58257;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1954.

KM Electron-transfer group; ETM; mismatch; genotyping;
KM gene expression; ss.

OS Synthetic

PN W0200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

XX

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

DR WPI; 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -

PS **Example 6**; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, CC monitoring gene expression.

SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match	Score	DB	Length
6.78;	121;	22;	936;

Matches 10; **Conservative** 484; **Mismatches** 299; **Indels** 0; **Gaps** 0;

[illegible]

Oy	1182	gggagatttgataatttttcaggatggtgcgaactaaacccagatgaagctca	1241
Db	183	242
Oy	1242	aaatacalcatatatagataltccaanaatctggaacctaaataatcaaaaagcgaagt	1301
Db	243	302
Oy	1302	agcagttcaatttctgaanaataatgattctccgccttcaatgaatgtcatttt	1361
Db	303	362
Oy	1362	gatagaagaggtacactcagtcgaatttgcgaactctgaacagtttaagttaat	1421
Db	363	422
Oy	1422	tccttgtgagatctgycacgaagacctggaactttggaccttcagaacattcttat	1481
Db	423	482
Oy	1482	acaaggaacaatcacatcacatgtgcactggtattgtcacccctccaaatagtgttla	1541
Db	483	542
Oy	1542	tgaccttaacatctgatgatggaacagagtaactagaagctatcctatgtactgaca	1601
Db	543	602
Oy	1602	aattctccagatccagatcagaatlttcgatgatgatagtaccttcagaaagtgtg	1661
Db	603	662
Oy	1662	atagatcatgatgatgtttgtccctccaggaataaaatltgatgtcatatccgtgttg	1721
Db	663	722
Oy	1722	aatgctcatcaagtcatacatatgtccaaatggaacagataatcaaatltgtacaga	1781
Db	723	782
Oy	1782	tttttgacaaccac	1794
Db	783	795

Search completed: April 9, 2002, 22:18:59
Job time: 7678 sec

Query Match

11.1%; Score 201; DB 22; Length 675;

Best Local Similarity 100.0%; Pred. No. 1.1e-45;

Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1616 ccagatcagaagtcctatgatgatgactcttcagaaaatgtgagataatcatgacat 1675
 |||
 Db 1 ccagatcagaagtcctatgatgatgactcttcagaaaatgtgagataatcatgacat 60
 Oy 1676 atgtttgtccctccagagataaaatgatcatatccgtgtgtggaatgtccatcaag 1735
 |||
 Db 61 atgtttgtccctccagagataaaatgatcatatccgtgtgtggaatgtccatcaag 120
 Oy 1736 tcatacatgtccacaatgagacagataatcaattgtcatcagattttgacaccaca 1795
 |||
 Db 121 tcatacatgtccacaatgagacagataatcaattgtcatcagattttgacaccaca 180
 Oy 1796 gttgcagaagatgtatctaa 1816
 |||
 Db 181 gttgcagaagatgtatctaa 201

RESULT 12

AAI39284

ID AAI39284 standard; DNA: 475 BP.

AC AAI39284;

DT 17-OCT-2001 (first entry)

DE Probe #7970 used to measure gene expression in human placenta sample.

KW Probe: microarray; human; placenta; antenatal diagnosis;

genetic disorder; ss.

OS Homo sapiens.

PN W0200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632386.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488897/53.

PT Human genome-derived single exon nucleic acid probes useful for

PI analyzing gene expression in human placenta

PS Claim 25; SEQ ID No 7970; 654pp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.

SQ Sequence 475 BP; 191 A; 70 C; 69 G; 145 T; 0 other;

Query Match

8.1%; Score 146.8; DB 22; Length 475;

Best Local Similarity 92.8%; Pred. No. 9.1e-31;

Matches 154; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Oy 1140 tcattgcccctaaatgctatttctgcaagaagttccatcatgagggatttgataat 1199
 |||
 Db 310 tcatttattgccaataattttaggcaagaagttccatcatgagggatttgataat 369
 Oy 1200 ttctcagatggttccaactaaacccagatgtccaagctacaataatcatatata 1259
 |||
 Db 370 ttctcagatggttccaactaaacccagatgtccaagctacaataatcatatata 429
 Oy 1260 ttcaaaaatcttgagccactaaaatcaaaaaggagaaagttagca 1305
 |||
 Db 430 ttcaaaaatcttgagccactaaaatcaaaaaggagaaagttagca 475

RESULT 13

AAFS8252

ID AAF58252 standard; DNA: 936 BP.

AC AAF58252;

DT 24-APR-2001 (first entry)

DE Oligonucleotide D1835.

KW Electron transfer group; ETM; mismatch; genotyping;

gene expression; ss.

OS Synthetic.

PN W0200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000; 2000WO-US20476.

PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.

PA (CLIN-) CLINICAL MICRO SENSORS INC.

PI Umek RM;

DR WPI; 2001-159728/16.

PT Nucleic acids containing electron-transfer group, useful as labels in

PI hybridization assays, e.g. for genotyping, allowing repeat analyses on

PS a single surface

CC Example 6; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.

SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match

6.7%; Score 121; DB 22; Length 936;

Best Local Similarity 1.3%; Pred. No. 1.7e-23;

Matches 10; Conservative 484; Mismatches 299; Indels 0; Gaps 0;

Oy 1002 tacaatactacagatcagatatttgagagagacacacatgtgcatcttgaaca 1061
 |||
 Db 3 tacaatactacagatcagatatttgagagagacacacatgtgcatcttgaaca 62
 Oy 1062 aaaagtcctcacaataccgcatccgagcaaaattggagtcataaagccagaagact 1121
 |||
 Db 63 aaaagtcctcacaataccgcatccgagcaaaattggagtcataaagccagaagact 122

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH16528 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AA892446 to
CC AA85893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

SQ Sequence 866 BP; 254 A; 162 C; 171 G; 274 T; 5 other;

Query Match	18.88	Score 342	DB 22	Length 866
-------------	-------	-----------	-------	------------

Best Local Similarity 91.28; Pred. No. 1.3e-84;

Matches 384; Conservative 0; Mismatches 30; Indels 7; Gaps 2;

OY	1	atgctcttggttcgaagacaacaattatataatataatgacccctcgaaatcaacttaaggtgtg	60
DB	439	atgtctcttggttcgaagacaacaattatataatataatgacccctcgaaatcaacttaaggtgtg	498
OY	61	acaattgcatgtgctatggtgtgtgttgaaagtctcttaagcccccatatacttaagcaaaaga	120
DB	499	acaattgcatgtgctatggtgtgtgttgaaagtctcttaagcccccatatacttaagcaaaaga	558
OY	121	actgaattatgtctcaagttgtgttaactatgtgtgaccagaacaaatgtgaaactaacttgcctg	180
DB	559	actgaattatgtctcaagttgtgttaactatgtgtgaccagaacaaatgtgaaactaacttgcctg	618
OY	181	ctctttaatggaacatgaagccctcccaataattataaaatgtagatatgtgtgc	240
DB	619	ctctttaatggaacatgaagccctcccaataattataaaatgtagatatgtgtgc	678
OY	241	tttcacagcgtgaagatcaaatatataaaagaagacccagggtatccaccagctgtgc	300
DB	679	tttcacagcgtgaagatcaaatatataaaagaagacccagggtatccaccagctgtgc	738
OY	301	tttgatcatttgaagtttggaaggaactttgggaacccctatcaactctgcacttgaagc	360
DB	739	tttgatcatttgaagttt--gaggggaactttgggaacccctatcaactctgcacttgaagc	796
OY	361	aagtaatttaacttaactgaactgaagacacaacaatgtglaagaagccttacgtgtgtggca	420
DB	797	agtaatttaactctact----gagcgcncaaatgtnagaagccttaaggtgttggct	851
OY	421	t 421	
DB	852	t 852	

RESULT 9

ID AAC28094 standard; cDNA; 336 BP

AC AAC28094;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 32169.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

05 Homo sapiens.

PN EP1033401-A2

06-SEP-2000 PD

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

PS Claim 1: SEQ ID 32169; 71pp + CD-ROM; English.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PI obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures

The present sequence is one of large number of 5' ESTs derived from cDNA libraries encoding secreted proteins. No ORF has yet been conclusively identified within the present sequence. The 5' ESTs were prepared from total human RNAs or polyA⁺ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (3'UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

Sequence 336 BP; 108 A; 70 C; 69 G; 86 T; 3 other,

Query Match	17.78;	Score 321.8;	DB 21;	Length 336;
-------------	--------	--------------	--------	-------------

Best Local Similarity 98.8%; Pred. No. 3.3e-79;

Matches 331; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Oy	758	caggaatcaacaacagttaagttaagttaattatcttaattgaagtaaccagtaagtc	817
Db	1	cagaaatacaacaacagtttaagtttaagtttaattcaatccaaagtaaccagtttaagtc	60
Oy	818	ggggatcaaggtctctgcagaaagtaactctgaatggatcaactgaaanaagatttag	877
Db	61	ggggatcagggtcctctgcagaaagtaactctgatctgatcaactgtaaaaagatttag	120
Oy	878	aatc-tgcaaatcttgacggccaatcagatcttgatgattatctgtcaatcaaaccttac	936
Db	121	aaaccttgaraatttgcagagccaatcagatcttgatgattatctgtcaatcgaacctgtac	180
Oy	937	gaacagcttccaagctctgtatcagatcatcatatcacgaagtagaaagatgtcaacgcta	996
Db	181	gacacgttccaagctctgtatcagatcatcatatcacgaagtagaaagatgtcaacgcta	240
Oy	997	ctgtgtacaatacttacagatcatcagattttgagaaggacacacatgtatgcattttg	1056
Db	241	ctgtgtacaatacttacagatcatcagattttgagaaggacacacatgtatgcattttg	300
Oy	1057	aaacaaaagctctctcaacaataccgcatccgagc	1091
Db	301	aaacaaaagctctctcaacaataccgcatccgagc	335

RESULT 10

ID AAH12625 standard; cDNA; 568 BP

AC AAH12625

QY 1505 -----gcactgggtatgtacccctccaatagtgttgtatgacctta 1550
DB 601 cttctgtgcagagaagcactgggtatgtacccctccaatagtgttgtatgacctta 660
QY 1551 cactgtatgtgagaagagactagaagcctatctgtatgttgcacaattcttc 1610
DB 661 cttctgtatgtgagaagagactagaagcctatctgtatgttgcacaattcttc 720
QY 1611 agattccacatcagaagctctgtatgtatgaccttcagaaaagtgtgatatacatca 1670
DB 721 agattccacatcagaagctctgtatgtatgaccttcagaaaagtgtgatatacatca 780
QY 1671 tggatattgttctgtcctccagagaataaattgatacatccgtgtgttggaatcttca 1730
DB 781 tggatattgttctgtcctccagagaataaattgatacatccgtgtgttggaatcttca 840
QY 1731 tcaagtcacatcagtcacaaatggacagatataatcttgcctgagattttggaca 1790
DB 841 tcaagtcacatcagtcacaaatggacagatataatcttgcctgagattttggaca 900
QY 1791 ccacagctgcagaagaatgtatctaa 1816
DB 901 ccacagctgcagaagaatgtatctaa 926

RESULT 6

AAH05303
ID AAH05303 standard; cDNA; 576 BP.

AAH05303;

26-JUN-2001 (first entry)

Human cDNA clone (5'-primer) SEQ ID NO:2138.

Human; primer: detection: diagnosis; antisense therapy: gene therapy; ss.

Homo sapiens.

EP:074617-A2.

07-FEB-2001.

28-JUL-2000: 2000EP-0116126.

29-JUL-1999: 99JP-0248036.

27-AUG-1999: 99JP-0300253.

11-JAN-2000: 2000JP-0118776.

02-JAN-2000: 2000JP-0183767.

09-JUN-2000: 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
MPI, 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -

Claim 1: SEQ ID 2138; 2537bp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

Sequence 576 BP: 167 A; 113 C; 114 G; 173 T; 9 other;

Query Match 30.0%; Score 545; DB 22; Length 576;
Best Local Similarity 98.7%; Pred. No. 8; 2e-141;
Matches 545; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 atgtcttggcttcacgaacaattatataataacccctgaatcaacttaaggtgt 60
DB 25 atgtcttggcttcacgaacaattatataataacccctgaatcaacttaaggtgt 84
QY 61 acaatgtcatgtctatgtgtgtggaagttcttaagccccaatataagcaaga 120
DB 85 acaatgtcatgtctatgtgtgtggaagttcttaagccccaatataagcaaga 144
QY 121 actgattattgtcctgaattgttaactatgtgagcagaacaatgttaactaactgtcgt 180
DB 145 actgattattgtcctgaattgttaactatgtgagcagaacaatgttaactaactgtcgt 204
QY 181 cttcttagtgaacacatgaagccttcacataattataaaatgtgatatgttcgc 240
DB 205 cttcttagtgaacacatgaagccttcacataattataaaatgtgatatgttcgc 264
QY 241 ttacacagctgaagaatcaaglatataaaagagagactcaaggtatcacagctctgcg 300
DB 265 ttacacagctgaagaatcaaglatataaaagagagactcaaggtatcacagctctgcg 324
QY 301 ttgacatcttggacgtttgaagggaacttggagccccaatataactcgaacttcaagc 360
DB 325 ttgacatcttggacgtttgaagggaacttggagccccaatataactcgaacttcaagc 384
QY 361 aagattttaactcactaactgaagaccacaatgtgaaagccttaactgtttggaca 420
DB 385 aagattttaactcactaactgaagaccacaatgtgaaagccttaactgtttggaca 444
QY 421 tctactcatatgtcaccgctctggacatataactatgtgtgagtttagccaagcg 480
DB 445 tctactcatatgtcaccgctctggacatataactatgtgtgagtttagccaagcg 504
QY 481 tatttgaacctgtcagctcctgtggcaagcagaagtgagcagcatcaattctt 540
DB 505 tatttgaacctgtcagctcctgtggcaagcagaagtgagcagcatcaattctt 564
QY 541 ctaaaaggtatgg 552
DB 565 ctaaaaggtatgg 576

RESULT 7

AAH07649
ID AAH07649 standard; cDNA; 854 BP.

AAH07649;

26-JUN-2001 (first entry)

Human cDNA clone (5'-primer) SEQ ID NO:4484.

DE
XX

PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

PS Claim 8: SEQ ID 12760: 2537pp + CD ROM: English.

XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 2077 BP: 657 A: 401 C: 393 G: 626 T: 0 other:

Query Match 94.5%; Score 1717; DB 22: Length 2077;
Best Local Similarity 95.3%; Pred. No. 0;

Matches 1816: Conservative 0; Mismatches 0; Indels 89; Gaps 1:

OY 1 atgctcttggtccagcaacaattatataataacccctgaactaagggtgt 60
DB 131 atgctcttggtccagcaacaattatataataacccctgaactaagggtgt 190
OY 61 acaattgtcaatgtcctatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120
DB 191 acaattgtcaatgtcctatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 250
OY 121 actgattttgtcagttgactatgtgtgtgtgtgtgtgtgtgtgtgtgt 180
DB 251 actgattttgtcagttgactatgtgtgtgtgtgtgtgtgtgtgtgtgt 310
OY 181 cctctttagtgaactatgaagccctccaataattataaaatgagagata 240
DB 311 cctctttagtgaactatgaagccctccaataattataaaatgagagata 370
OY 241 ttccagagctgaataatccaataataataaaatgagagata 300
DB 371 ttccagagctgaataatccaataataataaaatgagagata 430
OY 301 ttgcatccttgagcttgaaggaactttgagagccctcatatccctcgac 360
DB 431 ttgcatccttgagcttgaaggaactttgagagccctcatatccctcgac 490
OY 361 aagattttaactactactatgaagcacaataatgtagaagccttaagtt 420
DB 491 aagattttaactactactatgaagcacaataatgtagaagccttaagtt 550
OY 421 tctactcatatgacccgtcttgagcattactaaatgtgtgagtccaccc 480
DB 551 tctactcatatgacccgtcttgagcattactaaatgtgtgagtccaccc 610
OY 481 tatcttgacactgactgtcagctcttgaggcaagcagaagtgaagcagcat 540
DB 611 tatcttgacactgactgtcagctcttgaggcaagcagaagtgaagcagcat 670
OY 541 ctaaaagtatagtgatgacccagacacacattccatcttgagagttcta 600

DB 671 ctaaaagtatagtgatgacccagacacacattccatcttgagagttcta 730
OY 601 ctgtcttgaagtgatgaatgaatcactcagcgtacacaaatctgcacatag 660
DB 731 ctgtcttgaagtgatgaatgaatcactcagcgtacacaaatctgcacatag 790
OY 661 ttgactcagataacacatgatttgaagagatctcgaagttggaagcttct 720
DB 791 ttgactcagataacacatgatttgaagagatctcgaagttggaagcttct 850
OY 721 atctatagccttcataccacacatccaatgaatgaatcagaagatcagaat 780
DB 851 atctatagccttcataccacacatccaatgaatgaatcagaagatcagaat 910
OY 781 ttaaggttcatcttcatcagagaggtacccagttacaggtcgggaatcagg 840
DB 911 ttaaggttcatcttcatcagagaggtacccagttacaggtcgggaatcagg 970
OY 841 agtaactctgattgtgacactggaagaaagattagatctgcgaattgacag 900
DB 971 agtaactctgattgtgacactggaagaaagattagatctgcgaattgacag 1030
OY 901 cagcatcagatgtatctctgcaatcagaacctgacagacgttccaagctg 960
DB 1031 cagcatcagatgtatctctgcaatcagaacctgacagacgttccaagctg 1090
OY 961 gtatcatatagcagaagtagaagaatgtcaacagcctatctgtcaaatga 1020
DB 1091 gtatcatatagcagaagtagaagaatgtcaacagcctatctgtcaaatga 1150
OY 1021 cagtaatttggagagagacacacactatgtgcatlttgaacacaaagctc 1080
DB 1151 cagtaatttggagagagacacacactatgtgcatlttgaacacaaagctc 1210
OY 1081 cgcattcagagaacaaattgaggtcattatagcccgagaagactattcag 1140
DB 1211 cgcattcagagaacaaattgaggtcattatagcccgagaagactattcag 1270
OY 1141 catgtgccctaaatgtcatcttgcacaaagaatctcacacatgagggcgat 1200
DB 1271 catgtgccctaaatgtcatcttgcacaaagaatctcacacatgagggcgat 1330
OY 1201 ttcaagatgtgtcgaactaaacccccagatgtcgaactcacaatacata 1260
DB 1331 ttcaagatgtgtcgaactaaacccccagatgtcgaactcacaatacata 1390
OY 1261 tcaaaaatctgagccactaaatacaaaaagagcagaagaatgaggtcatt 1320
DB 1391 tcaaaaatctgagccactaaatacaaaaagagcagaagaatgaggtcatt 1450
OY 1321 aataatggtatctcccgcttccaatgaatgtcactttgtatagaagggtac 1380
DB 1451 aataatggtatctcccgcttccaatgaatgtcactttgtatagaagggtac 1510
OY 1381 agtgaatttgcacaacctgcgaacaaatgaatgaatgaatgaatgaatga 1440
DB 1511 agtgaatttgcacaacctgcgaacaaatgaatgaatgaatgaatgaatga 1570
OY 1441 gaagacctggaactttgagaccttccagcaccatttctatagaaggaata 1500
DB 1571 gaagacctggaactttgagaccttccagcaccatttctatagaaggaata 1630
OY 1501 tatg----- 1504
DB 1631 tatgattgataacagatgttctagttgagatcatalaacaatactaaat 1690
OY 1505 -----gacttggtattgtacccctcaatat 1531
DB 1691 aaaaacatcgatgattccttctctgtgagcagaacacactgggtattgata 1750
OY 1532 gtgttgtatgaccttatacacttgatgagtagaacagagtagtagaagctat 1591

Db 241 AACCCGAGTGTCAAGTCACAAATACATCATTTATATGATTCAAAAATCTGGACCACCTA 300
 Qy 1280 aaatcaaaaagagcagaaagta-gcagttcatttctgtgaasaataatggtatctccg 1338
 |||||
 |||||
 Db 301 AAAATCAAAAAGGACGANAATACGACGATTCATTTGTGAAAAATAATGGTATTTCTCCG 360
 Qy 1339 ctccaatgatcttacttctt-gatagaagagagacactcaagtaaatctgcaact 1397
Db 361 CTTTCAAATGGATGTCTACTTTTGGATAGGAGGGGGGTACACTCGTGGAAATTTGGCAAA 420			
Qy 1398 ct----cgacaagtttaatagttaatcc 1424			
 Db 421 CTTTTCGACCAAGTTTATTTGTGTTATTTTC 451

RESULT	14
AL520360	
LOCUS	782 bp mRNA
DEFINITION	AL520360 LRT1.NC004_NBC2 Homo sapiens cDNA clone CSDB006YH17 5'

ORGANISM	Homio sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 782)
AUTHORS	Li, M.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES	Location/Qualifiers
SOURCE	1..782

BASE COUNT	242 a	130 c	149 g	247 t	14 others
ORIGIN					

Query Match	20.18;	Score 365.4;	DB 10;	Length 782;
Best Local Similarity	96.28;	Pred. No. 4.4e-78;		
Matches 357;	Conservative 14;	Mismatches 0;	Indels 0;	Gaps 0

OY	1	agctcccttggcttcacacacaactatataatataaaccccttgaaacaaactcaagggctg	60
Db	412	atgctcttggcttcacacacaaattatataatamaccctcccttgatmaacttaagggtgtg	471
OY	61	acaattgtccaatgctatggtgtgtgaaagtcctttaagcccccatactcaagcaaga	120
Db	472	acaattgtccaatgctatggtgtgtgaaagtcctttaagcccccatactcaagcaaga	531
OY	121	actgatatatgtctcaagttgttaactatcttgagaccacagaacaatgtaaacctaacttcctg	180
Db	532	actgatatatgtctcaagttgttaactatcttgagaccacagaacaatgtaaacctaacttcctg	591

Qy	181	cttttttgtagtgaaccccttgaagccctcccaataattatataaaatggaatgtgttcg	240
Db	592	ctctttttatgtgaacctatggaagcccttmmatattttatataaaatgagattgttcgc	651
Qy	241	tttcacagagctgaagattccaaglatataaaaggagactcaggatcaccaagctctgac	300
Db	652	tttmmagcctgagagattmaagttatattaaaaagagamtmgaggtatcacacactctgcg	711
Qy	301	tttgcatccttgaagcttttggaggaaactttgggaacccctatcacactctgcacttaagc	360
Db	712	ttttgcattcttggagctttggaaggaaactttggagacmccctatcatacactctgcacttaagc	771
Qy	361	aagatatttaa	371
Db	772	aagatatttttaa	782

RESULT	15
BF696927	
LOCUS	882 bp mRNA EST 22-DEC-2000
DEFINITION	60212699.F1 NIH_MGC_56 Homo sapiens CDNA clone IMAGE:4286230 5', mRNA sequence.
ACCESSION	BF696927
VERSION	BF696927.1 GI:11982335
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

FEATURES	Location/Qualifiers
source	1. .882

BASE COUNT	259 a	194 c	203 g	226 t
ORIGIN				

Query Match	19.8%	Score 359.6;	DB 11;	Length 882;
Best Local Similarity	88.6%;	Pred. No. 1,le-76;		
Matches 459;	Conservative	0;	Mismatches 49;	Indels 10;
				Gaps 6;

07 330 ggagagccatcatcactcgcacatccaagatttttaacttcactcagtagaacca 389

BASE COUNT 259 a 164 c 166 g 259 t 6 others
 ORIGIN

Query Match 28.8%; Score 522.4; DB 10; Length 854;
 Best Local Similarity 99.3%; Pred. No. 4.3e-116;
 Matches 545; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

OY 1 atgtcttgggtccagcaacaattatataataacccctgaactaactaaggtgt 60
 |||||||
 Db 132 ATGCTTTGGTCCAGCACAAATATATATATACACCCCTGAATCACTTAAGGCTGT 191
 OY 61 acaattgcaatgctatgtgtgtgtgaagttctttaaagcccatatctaaagcaagga 120
 |||||||
 Db 192 ACAATTTGCAATGCTATGTTGTTGTTGCAAGTTCTTTTAAGCCCATATCTTAAGCAAGGA 251
 OY 121 actgattatgtcctcagttgaactatgttgaccagacaactgtaaaactaacttgctgt 180
 |||||||
 Db 252 ACTGATTAATGCTCAGTTGTAAGTATGTGACACCAATGTAACCTGCTGCTG 311
 OY 181 ctctttagtggaaacttgaaagccctcccaataattataaaatggagatattgtctgc 240
 |||||||
 Db 312 CTCTTTAGTGAACATATGAAGCCCTTCATATATTAATAAATGAGATATATGTTGCGC 371
 OY 241 ttccacagctggaagattcaaglatataaaagagagacaggtatcaccagctctgagc 300
 |||||||
 Db 372 TTTTCACAGGCTGAAGATTAAGTATATAAAGAGAGACTCAGGGTATACACAGCTCTGCG 431
 OY 301 ttggacatcttgacgtttgaggaactttggagagccctcatcaccctgcaactcaagc 360
 |||||||
 Db 432 TTTGCACTTTTGACGTTTGAGGAGACTTTTGAGCCCTTATACCTCCACCTTCAGC 491
 OY 361 aagattattcaactcaactaactagagaccacaataatgtagaagccttaagcttcttgagca 420
 |||||||
 Db 492 AAGTATTTTAACTTCACTACTAGAGACACCAAAATGTTAGAACCTTTAGCTGTTTGCGCA 551
 OY 421 tctactcaatgtgcaacgctcttgagcaacttaaaatgtgtgagcttcaagc -aatgca 479
 |||||||
 Db 552 TCTACTCAATGTGACCGCTTTTGACATTTCTAAATTTGTGATGTTCAGCCAAATGCA 611
 OY 480 gtaatttgaccactgtgacgtctcttgaggcaagagagatggagc-ggaagcctcatc 538
 |||||||
 Db 612 GTAATTTTGACCTACTTGTCACTTTGGCAAGACGAGAGTGGAGCGGNGCATTTTC 671
 OY 539 ttctaaag 547
 |||||||
 Db 672 TTCTAAAG 680

RESULT 11
 AV752694 529 bp mRNA EST 19-OCT-2000
 LOCUS AV752694 NPD Homo sapiens cDNA clone NPDAEB10 5', mRNA sequence.
 DEFINITION AV752694
 AV752694
 VERSION AV752694.1 GI:10910542
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 529)
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 Song, H., Peng, Y., Gu, Y., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N.,
 Qian, B., Liu, F., Ou, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu
 'M., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Ye, M., Zhang, O.
 ' Han, Z., Chen, Z., Hu, R. and Chen, J.
 Homo sapiens NPD library cDNA clones
 Unpublished (2000)
 Contact: Qinghua Zhang
 Shanghai Institute of Endocrinology, Rui-Jin Hospital
 197 Rui-Jin II Road, Shanghai 200025, P. R. China
 Tel: 86-21-64370045(ex.663332)

FEATURES
 source
 1..529
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NPDAEB10"
 /clone_1ib="NPD"
 /tissue_type="pituitary"
 /dev_stage="Adult"
 /lab_mosi="SOLR"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 168 a 106 c 104 g 150 t 1 others
 ORIGIN

Query Match 26.0%; Score 472.8; DB 10; Length 529;
 Best Local Similarity 95.4%; Pred. No. 4.2e-104;
 Matches 497; Conservative 0; Mismatches 23; Indels 1; Gaps 1;

OY 774 gtttaagtttagagtttcattcatctatgagaggtaccagttcaggtcggggaatcaggtctt 833
 |||||||
 Db 1 GTTAATTTAGAGTTTCACTTCATCTGAGGTACCACTTACGCTGCGGATCAGGCTCTT 60
 OY 834 gccagaagaactcctgatagtgagatcaactgaaagaatgtagaactgtcaaatgtgac 893
 |||||||
 Db 61 GCCAGAAAGTAACTCTGATGTGATCACTGAAAGAGATTAAGATTCGAAATTTGAC 120
 OY 894 agccaatcagagcttgagatgttatctgtcaatcaaacctcagcagcacttccaagtc 953
 |||||||
 Db 121 AGCCAATCGCAATTCAGATGTATCTGTCAATCAGAACTGACGACACTTTCCAAGCTC 180
 OY 954 tggatcagatcatcatacagagtagaagaatgtcaacagctatctgtacaatacttac 1013
 |||||||
 Db 181 TGGATCAGTATCATTTATACGAGTAGAAGATGTCAACAGTATCTGTACAAATCTTAC 240
 OY 1014 agatcatcagatatttggaagagacccactatgtgccattttgaaacaaagctctcca 1073
 |||||||
 Db 241 AGATCATCAGTATTTGGAGAGACACCACTATGTCCATTTTGAACAAAGCTCTCCA 300
 OY 1074 acaatcagcagctcagagcaaatltgagtcatalaagcccaagagacatttcaagtcgt 1133
 |||||||
 Db 301 ACAATACCGCATCCGAGCAAAATTTGAGTCAATATAAGCCCAAGAGACTATTTCAAGTCTG 360
 OY 1134 taaactcattgcccctaaatgtcatttgcctcagaagagttccacatagagcgagattga 1193
 |||||||
 Db 361 TAACTTCAATTTGGCTTAAATGTTCATTTGCTTCAGAAAGTT-CACATGAGGCGGNATTTGA 419
 OY 1194 tataattttcagagatgagcaactaaaccccaagatgtcagaagtcacagaataactcatt 1253
 |||||||
 Db 420 TATTAATTTTTCAGATGTGCTCCACTTAAACCCAGATGTTCAGCTACAAATATCATATT 479
 OY 1254 atatgattcaaaaactcggaccactaaataaaatcaaaaagagac 1294
 |||||||
 Db 480 ATATGATTTTAAAAATCTTGACCACTTAAATCGACCGGCG 520

RESULT 12
 AA351459/c 429 bp mRNA EST 21-APR-1997
 LOCUS AA351459 Infant brain Homo sapiens cDNA 5' end, mRNA sequence.
 DEFINITION AA351459
 AA351459
 VERSION AA351459.1 GI:2003779
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

```

/tissue_type="rhadomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/organism="Muscle; Vector: pOT7; Site:1: EcoRI;
Site:2: XhoI; cDNA made by oligo-dr priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      239 a      154 c      145 g      198 t
ORIGIN

```

```

Query Match      31.7%: Score 575.6; DB 10; Length 736;
Best Local Similarity 98.4%: Pred. No. 5.5e-129;
Matches 634; Conservative 0; Mismatches 4; Indels 6; Gaps 5;

```

```

Oy 727 agcccttaccacaaacttcaatcaatgaatcagagaacacgaatgaattgaag 786
    |||||||
Db 1 agccttaccacaaacttcaatcaatgaatcagagaacacgaatgaattgaag 60
Oy 787 ttacattctcagagaggaacaggaacaggaacaggaacaggaacaggaac 846
    |||||||
Db 61 ttacattctcagagaggaacaggaacaggaacaggaacaggaacaggaac 120
Oy 847 tctgagtgatgacacacacacacacacacacacacacacacacacacacac 906
    |||||||
Db 121 tctgagtgatgacacacacacacacacacacacacacacacacacacacac 180
Oy 907 tcaagatgacacacacacacacacacacacacacacacacacacacacac 966
    |||||||
Db 181 tcaagatgacacacacacacacacacacacacacacacacacacacacac 240
Oy 967 ttatacagagtagaagaatgacacacacacacacacacacacacacacacac 1026
    |||||||
Db 241 ttatacagagtagaagaatgacacacacacacacacacacacacacacacac 300
Oy 1027 ttgagagaggaacacacacacacacacacacacacacacacacacacacac 1086
    |||||||
Db 301 ttgagagaggaacacacacacacacacacacacacacacacacacacacac 360
Oy 1087 cgaagcaaatgagacacacacacacacacacacacacacacacacacacac 1146
    |||||||
Db 361 cgacacaaattgagacacacacacacacacacacacacacacacacacacac 420
Oy 1147 cctaaatgacacacacacacacacacacacacacacacacacacacacac 1206
    |||||||
Db 421 cctaaatgacacacacacacacacacacacacacacacacacacacacac 479
Oy 1207 gatgagcaaatgagacacacacacacacacacacacacacacacacacacac 1266
    |||||||
Db 480 gatgagcaaatgagacacacacacacacacacacacacacacacacacacac 539
Oy 1267 atctgagacacacacacacacacacacacacacacacacacacacacac 1326
    |||||||
Db 540 atctgagacacacacacacacacacacacacacacacacacacacacac 596
Oy 1327 ggaatttcccgcttcaatgaatgacacacacacacacacacacacacacac 1370
    |||||||
Db 597 ggaatttcccgcttcaatgaatgacacacacacacacacacacacacacac 638

```

```

RESULT 8
BF027455 920 bp mRNA EST 10-OCT-2000
LOCUS 601672751F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3955573 5'
DEFINITION mRNA sequence.
ACCESSION BF027455
VERSION BF027455.1 GI:10735167
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC/DCID/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LICM830 row: e column: 14
High quality sequence start: 52
High quality sequence stop: 746.
FEATURES
source
location/Qualifiers
1..920
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:395573"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/organism="Skin; Vector: pOT7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dr priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      275 a      189 c      189 g      267 t
ORIGIN

```

```

Query Match      30.2%: Score 548.6; DB 11; Length 920;
Best Local Similarity 92.0%: Pred. No. 1.9e-122;
Matches 702; Conservative 0; Mismatches 44; Indels 17; Gaps 11;

```

```

Oy 7 ttggtccagcaaatattatataataacacacacacacacacacacacacacac 66
    |||||||
Db 97 ttggtccagcaaatattatataataacacacacacacacacacacacacacac 155
Oy 67 gtcaatgacacacacacacacacacacacacacacacacacacacacacac 126
    |||||||
Db 156 gtcaatgacacacacacacacacacacacacacacacacacacacacacac 215
Oy 127 tatgtcaggttgacacacacacacacacacacacacacacacacacacacac 186
    |||||||
Db 216 tatgtcaggttgacacacacacacacacacacacacacacacacacacacac 275
Oy 187 agtgaacacacacacacacacacacacacacacacacacacacacacacac 246
    |||||||
Db 276 agtgaacacacacacacacacacacacacacacacacacacacacacacac 335
Oy 247 aggcgaagatgacacacacacacacacacacacacacacacacacacacac 306
    |||||||
Db 336 aggcgaagatgacacacacacacacacacacacacacacacacacacacac 395
Oy 307 tcttgacacacacacacacacacacacacacacacacacacacacacacac 365
    |||||||
Db 396 tcttgacacacacacacacacacacacacacacacacacacacacacacac 453
Oy 366 tttaacttcaactcagagacacacacacacacacacacacacacacacacacac 425
    |||||||
Db 454 tttaacttcaactcagagacacacacacacacacacacacacacacacacacac 512
Oy 426 tcaatgacacacacacacacacacacacacacacacacacacacacacacac 485
    |||||||
Db 513 tcaatgacacacacacacacacacacacacacacacacacacacacacacac 571

```


Email : filiang@lifetech.com URL :
 http://fulllength.invitrogen.com*
 BASE COUNT 290 a 176 c 166 g 256 t 1 others
 ORIGIN

Query Match 38.5%: Score 699; DB 10; Length 889;
 Best Local Similarity 89.8%: Pred. No. 7.3e-159;
 Matches 798; Conservative 1; Mismatches 1; Indels 89; Gaps 1;

```

Oy 871 gatttagaatcgcgaattgacagcccaatcagcattcagatgatatcgtcaatcaaa 930
    |||||
Db 1 GATTTAGAATCTGCCAAATTTGACCCCAATCGCATTCAGATGTTATCTGTCACAGAA 60
Oy 931 cctgacgacagctcttccagctcgtgacagtaataatatacagaggtagaagatgtcaa 990
    |||||
Db 61 CCTGACGACAGCTTTCACAGCTCTGATCAGTATCATATACGAGAGTGAAGAGTGCAC 120
Oy 991 cagctatcgtcctacatctcagatcatcagatcttggagaagacccctatgtcc 1050
    |||||
Db 121 CAGCTATCTGCTACAAATCTTACATATCATGTTTGGAGAGACACCCATATGTGCC 180
Oy 1051 atttgaacaaaagaagctcctcaacaataccgcatccgagcaaaattgagtcataa 1110
    |||||
Db 181 ATTTGAAACAAAAGCTCTCAACATACCGCATCCGACAAATGTGAGTCATATAG 240
Oy 1111 cccagaagactatctcagtcgttaactcattcgcctaaatgtcattgtcgaaga 1170
    |||||
Db 241 CCCAGAAAGCTATTTTCAGTCTGTTAACTTATTCCTTAAATGTCATTTGCTGCAGAA 300
Oy 1171 gtccacatgagggcgaatttgatataatcttccagagtggtgcaactaaacccagat 1230
    |||||
Db 301 GTTCCACATGAGGGCGGATTTGGATATATTTTTCAGATGCTGCACTAAACCCACAT 360
Oy 1231 gtcaagctacaaaatacatcatatagatccaanaatcgtgaccataaatacaaaa 1290
    |||||
Db 361 GTCAAGCTACAAAATACATCATATATGATTCAAAACCTGAGCACATAAAATCAAAA 420
Oy 1291 ggaacaaagtagcagctcattcttgtaaaaataatglatctccgcttcaaatgaa 1350
    |||||
Db 421 GGACGAAAGTAGCAGTTTCTTGTGAAAATAATGATTTCTCCCTTTCAAAATGAA 480
Oy 1351 tgtctactttgatalagaagagtagtaactcagtgaaatttgcaactcgaacagtt 1410
    |||||
Db 481 TGTCTACTTTGTAGTAGAAGGAGTAGTCACTAGTGAATTTGCAAACTCTGCAACAGTT 540
Oy 1411 aatagtgtaattcctgtgagatctggccagagaagacctggaaactttgacccctcaga 1470
    |||||
Db 541 AATAGTGAATTCCTGAGATCTGGCCACGAAAGCTGGAACCTTTTGACCTTTTCAGCA 600
Oy 1471 ccatttctatacagaagacaatacatcactatg----- 1504
    |||||
Db 601 CCATTTCTTATACAAAGAAATACATCATGTGATGTAACAGTGTTCAGTTTGAGA 660
Oy 1505 ----- 1504
    |||||
Db 661 TCATATACAAATCTAAATTCCTGTTGATMAAACATCGTGATTCCTTCTGTGGA 720
Oy 1505 ---gcactgggattgtacccctcacaatagtgtgtgtgtagaccttaacactgagat 1561
    |||||
Db 721 GAACTACTGGGATATTGTAACCCCTCCAAATGTTGTTGTATGACCTTTACCTTGATGAT 780
Oy 1562 ggaacagagtagactagaagcctatcactatgagatctgacaactcttccaactcagca 1621
    |||||
Db 781 GGAACAGAGTAGTAGAAGCCTATCTCATGATTTCTGACAAATCTTCCAGATTTCCAGCA 840
Oy 1622 tcagaagttctgtagatgtagtaccctcagaagaagtgtgatalgatac 1670
    |||||
Db 841 TCAGAAAGTTCGATGATGATGACCTTCAGAAAGTGTGRTATGATCA 889
  
```

RESULT 2
 BE891252

```

LOCUS BE891252 926 bp mRNA EST 20-OCT-2000
DEFINITION 60143077F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3917299 5',
mRNA sequence.
ACCESSION BE891252
VERSION BE891252.1 GI:10350399
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 926)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgabs-remail.nih.gov
Tissue Procurement: ATCC/DC/DTP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8743 row: 3 column: 20
High quality sequence stop: 636.
FEATURES
Location/Qualifiers
source 1..926
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3917299"
/clone_11b="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: skin; Vector: PCMV-SPOrt6; Site: 1; Note:
Site: 2; SalI: Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 271 a 205 c 212 g 238 t
ORIGIN
Query Match 33.7%: Score 611.2; DB 11; Length 926;
Best Local Similarity 94.0%: Pred. No. 1.3e-137;
Matches 679; Conservative 0; Mismatches 38; Indels 5; Gaps 4;
Oy 87 gaagttcttaagccccaatctaaagcaagaactgatatgtcagttgtaactat 146
    |||||
Db 1 GAGTTCCTTTAAGCCCCCATATCTAAGCAAGAAAGAGATGATTGCTCAGTGTACTAT 60
Oy 147 tgtgaccagacaagaatgtaaaactaacttgcctgctcttaagtggaactatgaagccct 206
    |||||
Db 61 TGTGGACCAACAAATGTAAACTTAAGCTTGGCTCTTTAGTGAAGTGAAGCCCT 120
Oy 207 tccaataattataaagtagagatatgttcgcttccacagagtgaaatcaagtata 266
    |||||
Db 121 TCATATATTTATTAATAAGTAGAGATATGTGCTTTCACAGGCTTAACATTCAGTATA 180
Oy 267 taaaagagaactcagggtatcacagcctcgtgcttgaacttgaagttgagggaac 326
    |||||
Db 181 TAAAGAGAGACTCAGGATATCACACAGCTCTGCTTGTGATCTTTGACGTTGAGGGAAC 240
Oy 327 ttggagccccaatacactcgcactcgaactcaagaagatttaacttaactcagaaga 386
    |||||
Db 241 TTGGAGGCCCTTATCTACTACTGCACTTCAACAGATATTTTAACCTTACTAGTAGGA 300
Oy 387 ccacaaaatgtagaagccttactggtttgggcatcactacataltcaccgtcttgagc 446
    |||||
Db 301 CCACAAATAGTGAAGCCTTAGTGTTGGGCGATCTACTCATATGTACCGCTTTGGAC 360
Oy 447 attactaaatgtgtgtagtgcagccaatgcagtalattgacgtgactgttcagctct 506
    |||||
Db 361 ATTACTAAATGTGTGATGTTGAGCAATGACGATATTTTGACGCTGACTGTCAGCTTT 420
  
```



```

: APPLICANT: Muller-Cohn, Judy
: APPLICANT: Stamp, Lisa
: TITLE OF INVENTION: No. 6204435e1 Pesticidal Toxins and Nucleotide
: TITLE OF INVENTION: Sequences Which Encode These Toxins
: NUMBER OF SEQUENCES: 134
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: US
: ZIP: 32606-6669
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/960,780
: FILING DATE: 30-OCT-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/029,848
: FILING DATE: 30-OCT-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanchik, David R.
: REGISTRATION NUMBER: 31,794
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 352-375-8100
: TELEFAX: 352-372-5800
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1045 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: INDIVIDUAL ISOLATE: 31J2
:
: US-08-960-780-20
:
Query Match 1.88; Score 33.2; DB 4; Length 1045;
Best Local Similarity 49.48; Pred. No. 5.2;
Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
:
QY 1285 caaaaggagcgaagtagcagttcatttggtaaaataatggtatctccgcttca 1344
: ||||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 245 CAAAAGGCCAAAGAAACAAGTGTTCATTAGAAAAAGATAATTAGTCCCATCAA 304
:
QY 1345 aatgaatgctactcttggatgaagaggtacactcagtgaaatttgcgaactctgaac 1404
: ||||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 305 ATTGAATATCAATCTGATTAAGCGTTAAACCCAGATAGTCAAAATGTTAAAGAAATTGAAA 364
:
QY 1405 aagttataatgtaattcctgtagatctgcccagaagaccctggaaactttg 1458
: ||||| | | | | ||| | ||| | ||| | ||| | ||| | ||| |
Db 365 TTATTTAATAATATAGTCAAAAACAAATCTCAGCAAGTGCACAAGCAAGAAATTG 418
:
RESULT 11
: Sequence 20, Application US/09073898
: Patent No. 6242669
: GENERAL INFORMATION:
: APPLICANT: Feltelson, Jerald S.
: APPLICANT: Schnepf, H. Ernest
: APPLICANT: Naraya, Kenneth E.
: APPLICANT: Stockhoff, Brian A.
: APPLICANT: Schmeits, James
: APPLICANT: Loewer, David
: APPLICANT: Dullum, Charles Joseph
: APPLICANT: Muller-Cohn, Judy
```

```

: APPLICANT: Stamp, Lisa
: APPLICANT: Morrill, George
: APPLICANT: Finstad-Lee, Stacey
: TITLE OF INVENTION: No. 624269e1 Pesticidal Toxins and Nucleotide
: TITLE OF INVENTION: Sequences Which Encode These Toxins
: NUMBER OF SEQUENCES: 144
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: US
: ZIP: 32606-6669
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/073,898
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/029,848
: FILING DATE: 30-OCT-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Sanders, Jay M.
: REGISTRATION NUMBER: 39,355
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 352-375-8100
: TELEFAX: 352-372-5800
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1045 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: INDIVIDUAL ISOLATE: 31J2
:
: US-09-073-898-20
:
Query Match 1.88; Score 33.2; DB 4; Length 1045;
Best Local Similarity 49.48; Pred. No. 5.2;
Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
:
QY 1285 caaaaggagcgaagtagcagttcatttggtaaaataatggtatctccgcttca 1344
: ||||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 245 CAAAAGGCCAAAGAAACAAGTGTTCATTAGAAAAAGATAATTAGTCCCATCAA 304
:
QY 1345 aatgaatgctactcttggatgaagaggtacactcagtgaaatttgcgaactctgaac 1404
: ||||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 305 ATTGAATATCAATCTGATTAAGCGTTAAACCCAGATAGTCAAAATGTTAAAGAAATTGAAA 364
:
QY 1405 aagttataatgtaattcctgtagatctgcccagaagaccctggaaactttg 1458
: ||||| | | | | ||| | ||| | ||| | ||| | ||| | ||| |
Db 365 TTATTTAATAATATAGTCAAAAACAAATCTCAGCAAGTGCACAAGCAAGAAATTG 418
:
RESULT 12
: Sequence 72, Application US/08220151
: Patent No. 5529780
: GENERAL INFORMATION:
: APPLICANT: Paolelli, Enzo
: APPLICANT: Limbach, Keith J.
: TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
: TITLE OF INVENTION: CANINE HERPESVIRUS gb, gc AND gd AND USES THEREFOR
```

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Ave., N. W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,325
FILING DATE: 2-MAY-1997
CLASSIFICATION: 514
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/14251
FILING DATE: 02-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/482,401
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/334,029
FILING DATE: 02-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0917,0240002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 6371 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: DNA (genomic)
US-08-836-325-13

Query Match
Best Local Similarity 1.9%; Score 33.8; DB 3; Length 6371;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 379 actgagaccacaagaatgtagaagccttgcgtgttggtggcgcctcctatgacgcg 438
DB 2320 ACTGAGCAATTCAAAATGACTTGCATAGGAATTTGCTCTTACTGGAATCTTTGCA 2379
QY 439 tcttgacattactaaatgtgtgacgtgtcagccaatgcagatatttgacctgactgt 498
DB 2380 GCTGAATGCTATTAAACATGATGCCATGATCATGATGATTTCCAGTAGGCTGG 2439
QY 499 cagctcttgagca 511
DB 2440 AATATTTTGACA 2452

RESULT 7
US-08-836-325-14
Sequence 14, Application US/08836325
Patent No. 6110672
GENERAL INFORMATION:
APPLICANT: Mandel, Gail
APPLICANT: Halegoua, Simon
APPLICANT: Borden, Laurence A.
TITLE OF INVENTION: Peripheral Nervous System Specific
TITLE OF INVENTION: Sodium Channels, DNA Encoding Therefor, Crystallization,
TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational
TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Using
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Ave., N. W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,325
FILING DATE: 2-MAY-1997
CLASSIFICATION: 514
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/14251
FILING DATE: 02-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/482,401
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/334,029
FILING DATE: 02-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0917,0240002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 6404 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: DNA (genomic)
US-08-836-325-14

Query Match
Best Local Similarity 1.9%; Score 33.8; DB 3; Length 6404;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 379 actgagaccacaagaatgtagaagccttgcgtgttggtggcgcctcctatgacgcg 438
DB 2353 ACTGAGCAATTCAAAATGACTTGCATAGGAATTTGCTCTTACTGGAATCTTTGCA 2412
QY 439 tcttgacattactaaatgtgtgacgtgtcagccaatgcagatatttgacctgactgt 498
DB 2413 GCTGAATGCTATTAAACATGATGCCATGATGATCATGATGATTTCCAGTAGGCTGG 2472
QY 499 cagctcttgagca 511
DB 2473 AATATTTTGACA 2485

RESULT 8
US-08-960-780-24
Sequence 24, Application US/08960780
Patent No. 6204435
GENERAL INFORMATION:
APPLICANT: Fellelson, Jerald S.
APPLICANT: Schepf, H. Ernest
APPLICANT: Narva, Kenneth E.
APPLICANT: Stockhoff, Brian A.
APPLICANT: Schneits, James
APPLICANT: Loewer, David
APPLICANT: Dullum, Charles Joseph
APPLICANT: Muller-Cohn, Judy
APPLICANT: Stamp, Lisa
TITLE OF INVENTION: No. 6204435el Pesticidal Toxins and Nucleotide

Query Match	1.9%	Score 34.8;	DB 4;	Length 1046;
Best Local Similarity	50.0%;	Pred. No. 1.8;		

OY 269 TSYGRIAYLE---SNSDVOLK-----DLESANLTANOHSVICOSEPDSPFGVCS 320
DB 417 TRTNGPLVTEELHSLSFETOLCOPGLVLETTSLPV-----VVI-----SNVS 462
OY 321 LRPGMSSVARSRLIA 337
DB 463 QLPSCMASILMYNMLVA 479

RESULT 14
US-08-852-091-4

; Sequence 4, Application US/08852091
; Patent No. 5883228
; GENERAL INFORMATION:
; APPLICANT: James E. Darnell, Jr.
; APPLICANT: Zilong Wen
; APPLICANT: Curt M. Horvath
; APPLICANT: Zhong Zhong
; TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL
; TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,091
; FILING DATE: 06-MAY-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/369,796
; FILING DATE: 06-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 750 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-852-091-4

Query Match 4.7%; Score 84; DB 2; Length 750;
Best Local Similarity 23.9%; Pred. No. 3.8;
Matches 47; Conservative 28; Mismatches 62; Indels 60; Gaps 10;

OY 183 VMDGTRTPPSMRVLIOD-----LVLEGDLSHIRLO-----N 215
DB 301 LMDRT---FSLFOQLIOSSFVERQPCMPHPRPLVLTGVOFTYKRLRLVYLQELNIN 357
OY 216 LTTDLIVTNRHVARSLKAGSFLRLYSLTHTKLQSNSENQTMLEF-HL-----HGG 268
DB 358 LKXKVL-FDQDVNERTIVGFRFNILGITHKVMNMEESTNGSLAEFRHLQLEKQKNG 416
OY 269 TSYGRIAYLE---SNSDVOLK-----DLESANLTANOHSVICOSEPDSPFGVCS 320
DB 417 TRTNGPLVTEELHSLSFETOLCOPGLVLETTSLPV-----VVI-----SNVS 462

OY 321 LRPGMSSVARSRLIA 337
DB 463 QLPSCMASILMYNMLVA 479

RESULT 15
US-08-820-754-4

; Sequence 4, Application US/08820754
; Patent No. 5976835
; GENERAL INFORMATION:
; APPLICANT: Darnell Jr., James E.
; APPLICANT: Schindler, Christian W.
; APPLICANT: Fu, Xian-Yuan
; APPLICANT: Wen, Zilong
; APPLICANT: Zhong, Zhong
; TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
; TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,754
; FILING DATE: 19-MAR-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,185
; FILING DATE: 11-MAR-1994
; APPLICATION NUMBER: US 07/980,498
; FILING DATE: 23-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/854,296
; FILING DATE: 19-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO US93/02569
; FILING DATE: 19-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/126,588
; FILING DATE: 24-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-073 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 750 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-820-754-4

Query Match 4.7%; Score 84; DB 2; Length 750;
Best Local Similarity 23.9%; Pred. No. 3.8;
Matches 47; Conservative 28; Mismatches 62; Indels 60; Gaps 10;

OY 183 VMDGTRTPPSMRVLIOD-----LVLEGDLSHIRLO-----N 215
DB 301 LMDRT---FSLFOQLIOSSFVERQPCMPHPRPLVLTGVOFTYKRLRLVYLQELNIN 357

```

RESULT 11
US-08-276-099A-12
: Sequence 12, Application US/08276099A
: Patent No.:5591825
:
GENERAL INFORMATION:
APPLICANT: McRalghe, Steven L
APPLICANT: Hou, Jinzhao
TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
TITLE OF INVENTION: BINDING ASSAYS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarradero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,099A
FILING DATE: 15-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59451-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
FILER: 910 277299
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 740 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-276-099A-12

```



```

Db 247 CNGLSMHQVYKRC-----IVGFHLK-----270
QY 104 LFEETGLGAPIIPRTSSKYFNFTEDHKVNEALR-----VNAS-----141
Db 271 TLLEGCCGEVYHVESEHGFRLDVESSHAIIOVYKLLQELLDWKIMNAEGCWETLLAA 330
QY 142 -----THMSPSTLLKCDVQPMQYFDLTQLLGK-----AE 173
Db 331 LEVLIRVEHHQOQFNITKOLNNAHVHFLTCOVLOEHRREGOLTSMPREVCSEFKIIAE 390
QY 174 VDMA-----SFLKLVWDGTRT-----189
Db 391 VLGSPPDELLLVLEFNEFLAVHPRTNVCHNPTFTSLHIIDGKIIOEKVQSLAYLRHS 450
QY 190 -----PEFSKVL-----ODLVLEGDLSHIHLONLTIDILYDNHNVH 228
Db 451 SSGGAFPSGFLVISPFAFTAAPPEGTSSNIVPQMAQMVNSRLP-----AFPTLTP 506
QY 229 VASLKVGSFLRIYSLHFKIOSM-----NSENOTMLSEFLHGGTSGYGCIVLPESNSDV 285
Db 507 LIRAKTLAASIG-FSV-DKIONIADANPEKONLGRPYALK--TSKEAFIISCESAKTV 562
QY 286 DOLKKELESANLTAN-----OHSDVICOSEPDSPFNGVSL--RP---PGW 326
Db 563 CEMEA-LIGAHASANGVSRGSPRFPRAKVHDKDVGTERSDDSPGDESTIPRPDNLKGL 621
QY 327 SSVARSRLIAS 338
Db 622 ASFORSOSTVAS 633

```

RESULT 2

```

US-09-021-323-1
: Sequence 1, Application US/09021323
: Patent No. 5929033
: GENERAL INFORMATION:
: APPLICANT: Tang, Y. Tom
: APPLICANT: Corley, Neil C.
: TITLE OF INVENTION: EXTRACELLULAR MUCOUS MATRIX
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Dr.
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/021,323
: FILING DATE: Filed Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PR-0477 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 510 amino acids
: TYPE: amino acid

```

```

: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: SINTNOT13
: CLONE: 1805338
US-09-021-323-1

```

```

Query Match 4.88; Score 85; DB 4; Length 510;
Best Local Similarity 20.7%; Pred. No. 1.6;
Matches 69; Conservative 48; Mismatches 121; Indels 96; Gaps 16;

```

```

QY 30 KFFRPYLSKTDYCSVVTIVDQNVKLTCLLSGNEALPIYKNGDIVRHLQIYV 89
Db 112 KFEKE--LSKREYVOLISYVEKLLNLTVRI--DIMENDISTYELD---FELIRVEVA 164
QY 90 KKETOGITSSGFALEPTEGLGAPIIPRTSSKYFNFT-----TEDHKVNEALRMAST 142
Db 165 EMERLVIO-----LKESEFGS--SEIVDQLEVERLNMHTLVLEKLETLDKNNVLAIR-----R 214
QY 143 HMSPSWTLKLCVQPMQYFDL-----TCOLLGKAEVD-----GASFLLKW 184
Db 215 EIVALKTKLKECEASKDQNTPVVHPPTPCSCGHGVNISKPSVYQLMMRGFSYLYGAM 274
QY 185 DGRTP-FPS---WRVLIOVLLEGDLSHIHLONLTIDILYDNHNVASLKV-----235
Db 275 GROYSPQHPNKGKLYW---VAPLNTDGRLELYRLYNTLDDLLYLN---ARELRITYGQ 327
QY 236 GSFRLIYSLHFKIOSMENSENOTMLSEFLHGGTSGYGCIVLPESNSDVQDKKOLESA 295
Db 328 GSGTAVYNNMNVYMYNTGN-----IARY 351
QY 296 NLTANQHSVDVICOSEPDSPFNGVSLRPPGWSV 329
Db 352 NLTTN--TIAVYOTLPMAVNNRFSYANVAMQDI 383

```

RESULT 3

```

US-09-308-375-2
: Sequence 2, Application US/09308375
: Patent No. 6300117
: GENERAL INFORMATION:
: APPLICANT: Genencor International, Inc.
: TITLE OF INVENTION: Proteases from Gram-Positive Organisms
: FILE REFERENCE: GC394-PCT
: CURRENT APPLICATION NUMBER: US/09/308,375
: EARLIER FILING DATE: 1993-05-14
: EARLIER APPLICATION NUMBER: EP9719636.4
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 2285
: TYPE: PRT
: ORGANISM: Bacillus subtilis
US-09-308-375-2

```

```

Query Match 4.88; Score 85; DB 4; Length 2285;
Best Local Similarity 19.88; Pred. No. 18;
Matches 70; Conservative 47; Mismatches 114; Indels 122; Gaps 17;

```

```

QY 1 MSLVPAT-----NYITPLNLQKGTIVNVYGVYKFFKPPYLSKTDYCSVVTIVDQIN 54
Db 1464 MQLMHPATAKSLGVNNAYDPVGNVAGT-----KYLAQOLE-----1498
QY 55 VKLCLFLSGNVEALPIYK--NGDIVRHL-----KIOV-YKRETOGITSSG 100
Db 1499 -----KFGNVEKALAAYNAGPQNVIAKYGIPPFKEQYVYKIMANYKSLSSATSSI 1552
QY 101 FASLFEETGLGAPIIPRTSSKYFNFTEDHKVNEALRMASTHMSPSWTLKLCVQPMQ 160
Db 1553 ASYITNNSAF-----RVSSKY-----GOESGLR--SSPHKGTDFAKAGTAIKSLD 1597

```

